SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Takanori OKURA
               Kakuji TORIGOE
               Masahi KURIMOTO

    (ii) TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE OF INDUCING THE PRODUCTION OF INTERFERON-γ

    (iii) NUMBER OF SEQUENCES: 35

    (iv) CORRESPONDENCE ADDRESS:
        (A) ADDRESSEE: BROWDY AND NEIMARK
        (B) STREET: 419 Seventh Street, N.W., Suite 300
        (C) CITY: Washington
        (D) STATE: D.C.
        (E) COUNTRY: USA
        (F) ZIP: 20004

    (v) COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: Patent In Release #1.0, Version #1.30

    (vii) PRIOR APPLICATION DATA:
        (A) APPLICATION NUMBER: JP 185,305/96
        (B) FILING DATE: 27-JUN-1996

    (viii) ATTORNEY/AGENT INFORMATION:
        (A) NAME: BROWDY, Roger L.
        (B) REGISTRATION NUMBER: 25,618
        (C) REFERENCE/DOCKET NUMBER: OKURA=1

    (ix) TELECOMMUNICATION INFORMATION:
        (A) TELEPHONE: 202-628-5197
        (B) TELEFAX: 202-737-3528

(2) INFORMATION FOR SEQ ID NO: 1:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 157 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn
1               5                  10                  15
Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp
            20                  25                  30
Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg Thr Ile Phe Ile
        35                  40                  45
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
    50                  55                  60
Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys Glu Asn Lys Ile
65                  70                  75                  80
Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys
                85                  90                  95
Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys
            100                 105                 110
Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu
            115                 120                 125
```

```
Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu
    130                     135                 140
Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO: 2:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 1120 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA to mRNA

        (iii) HYPOTHETICAL: No

        (iv) ANTI-SENSE: No

        (vi) ORIGINAL SOURCE:
            (A)ORGANISM: human
            (F)TISSUE TYPE: liver

        (iX) FEATURE:
            (A) NAME/KEY: 5'UTR
            (B) LOCATION: 1..177
            (C) IDENTIFICATION METHODS: E
            (A) NAME/KEY: leader peptide
            (B) LOCATION: 178..285
            (C) IDENTIFICATION METHODS: S
            (A) NAME/KEY: mat peptide
            (B) LOCATION: 286..756
            (C) IDENTIFICATION METHODS: S
            (A) NAME/KEY: 3'UTR
            (B) LOCATION: 757..1120
            (C) IDENTIFICATION METHODS: E

        (xi)SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
GCCTGGACAG TCAGCAAGGA ATTGTCTCCC AGTGCATTTT GCCCTCCTGG CTGCCAACTC    60
TGGCTGCTAA AGCGGCTGCC ACCTGCTGCA GTCTACACAG CTTCGGGAAG AGGAAAGGAA   120
CCTCAGACCT TCCAGATCGC TTCCTCTCGC AACAAACTAT TTGTCGCAGG AATAAAG      177
ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG    225
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met
    -35                     -30                 -25
AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAT GAA AAC    273
Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn
-20                     -15                 -10                 -5
CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA    321
Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile
                1                   5                   10
AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT    369
Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro
                15                  20                  25
CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA GAT AAT GCA CCC CGG    417
Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg
    30                      35                  40
ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG    465
Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met
45                      50                  55                  60
GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT    513
Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys
                65                  70                  75
GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC    561
Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile
                80                  85                  90
AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA AGT GTC CCA GGA    609
```
```
                            - 31 -
```

```
Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly
     95                  100                 105


CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT    657
His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe
    110                 115                 120
CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA    705
Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys
125                 130                 135                 140
GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA    753
Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu
            145                 150                 155
GAC TAGCTATTAA AATTTCATGC CGGGCGCAGT GGCTCACGCC TGTAATCCCA         806
Asp
GCCCTTTGGG AGGCTGAGGC GGGCAGATCA CCAGAGGTCA GGTGTTCAAG ACCAGCCTGA   866
CCAACATGGT GAAACCTCAT CTCTACTAAA AATACTAAAA ATTAGCTGAG TGTAGTGACG   926
CATGCCCTCA ATCCAGCTA CTCAAGAGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG   986
GTAGAGGTTG TGGTGAGCCG AGATTGCACC ATTGCGCTCT AGCCTGGGCA ACAACAGCAA  1046
AACTCCATCT CAAAAAATAA AATAAATAAA TAAACAAATA AAAAATTCAT AATGTGAAAA  1106
AAAAAAAAAA AAAA                                                   1120
```

(2) INFORMATION FOR SEQ ID NO: 3:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 135 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: Genomic DNA

        (vi) ORIGINAL SOURCE:
            (A) ORGANISM: human
            (F) TISSUE TYPE: placenta

        (iX) FEATURE:
            (A) NAME/KEY: exon
            (B) LOCATION: 1..135
            (C) IDENTIFICATION METHODS: S

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
 AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA    47
 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
  -5                  1                   5                  10
GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT    95
Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn
                15                  20                  25
CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G             135
Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp
            30                  35                  40
```

(2) INFORMATION FOR SEQ ID NO: 4:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 134 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: Genomic DNA

        (vi) ORIGINAL SOURCE:
            (A) ORGANISM: human
            (F) TISSUE TYPE: placenta

(iX) FEATURE:
       (A) NAME/KEY: exon
       (B) LOCATION: 1..134
       (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
 AT AAT GCA CCC CGG ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC        47
Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser
40                  45                  50                  55
CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT        95
Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile
                60                  65                  70
TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT TCC TTT AAG                   134
Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys
            80                  85
```

(2) INFORMATION FOR SEQ ID NO: 5:

       (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 87 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: Genomic DNA

       (vi) ORIGINAL SOURCE:
            (A) ORGANISM: human
            (F) TISSUE TYPE: placenta

       (iX) FEATURE:
            (A) NAME/KEY: exon
            (B) LOCATION: 1..87
            (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
GAATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG       50
        Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val
        -35                 -30                 -25
GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G                      87
Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
    -20                 -15                 -10
```

(2) INFORMATION FOR SEQ ID NO:6:

       (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 12 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: Genomic DNA

       (vi) ORIGINAL SOURCE:
            (A) ORGANISM: human
            (F) TISSUE TYPE: placenta

       (iX) FEATURE:
            (A) NAME/KEY: exon
            (B) LOCATION: 1..87
            (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
CT GAA GAT GAT G
Ala Glu Asp Asp Glu
-10
```

(2) INFORMATION FOR SEQ ID NO: 7:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 2167 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: Genomic DNA

        (vi) ORIGINAL SOURCE:
            (A) ORGANISM: human
            (F) TISSUE TYPE: placenta

        (iX) FEATURE:
            (A) NAME/KEY: exon + 3'UTR
            (B) LOCATION: 1..2167
            (C) IDENTIFICATION METHODS: E

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA    48
Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile
85                  90                  95                  100
TTC TTT CAG AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA    96
Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu
                    105                 110                 115
TCT TCA TCA TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC    144
Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp
                120                 125                 130
CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT    192
Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser
            135                 140                 145
ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTTCA TGCCGGGCGC    246
Ile Met Phe Thr Val Gln Asn Glu Asp
        150                 155
AGTGGCTCAC GCCTGTAATC CCAGCCTTT GGGAGGCTGA GGCGGGCAGA TCACCAGAGG    306
TCAGGTGTTC AAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA    366
AAAATTAGCT GAGTGTAGTG ACCATGCCC TCAATCCCAG CTACTCAAGA GGCTGAGGCA    426
GGAGAATCAC TTGCACTCCG GAGGTGGAGG TTGTGGTGAG CCGAGATTGC ACCATTGCGC    486
TCTAGCCTGG GCAACAACAG CAAAACTCCA TCTCAAAAAA TAAAATAAAT AAATAAACAA    546
ATAAAAAATT CATAATGTGA ACTGTCTGAA TTTTTATGTT TAGAAAGATT ATGAGATTAT    606
TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCAGT CTTGAAAAAC ATCATTAAGA    666
AATGAATGAA CTTTCACAAA AGCAAACAAA CAGACTTTCC CTTATTTAAG TGAATAAAAT    726
AAAATAAAAT AAAATAATGT TTAAAAAATT CATAGTTTGA AAACATTCTA CATTGTTAAT    786
TGGCATATTA ATTATACTTA ATATAATTAT TTTTAAATCT TTTGGGTTAT TAGTCCTAAT    846
GACAAAAGAT ATTGATATTT GAACTTTCTA ATTTTTAAGA ATATCGTTAA ACCATCAATA    906
TTTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAAT TTCCTCCAAA GTCAGTATAT    966
TTTTAAAATT CAGTTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC    1026
CAACTGACAT TCATCCTAGG AATGCAAAGA TGGTTTAATA TCCTAAAATC AATTAACATA    1086
ACATACTATA TTAATAAAGT ATCAAAACAG TATTCTCATC TTTTTTTCTT TTTTCACAAT    1146
TCCTTGGTTA CACTATCATC TCAATAGATG CAGAAAAAGC ATTTGACAAA ATCCAATTCA    1206
TAATAAAAAT TCTCAAACTT GAAAGAGAAT ATCATAAAGG CATCTATGAA AAACCTACAG    1266
CTAATATCAT ACTTAACGAT GAAAAACTGA ATTATTTTAC CCTAAGATCA AGAATAATGC    1326
AAGCATGTCA GCTCTTGCAA CTTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA    1386
CCATACAATA AATAAAAATA AAAGGCACCC AGATTAGAAA GGAAGTCTTT ATTTGCAGAC    1446
AACATGGTTC TTTATGCAGA AAACCGTCAG GAATACACAC ACATGTTAGA ACTAATAAGT    1506
TCAGCAAGGT TGCAGGTTGC AATATCAATA TGCAAAAATA CATTGAAGGC TGGGCTCAGT    1566
GGAGATGGCA TGTACCTTTC GTCCCAGCTA CTTGGGAGGC TGAGGTAGGA GGATCACTTG    1626
AGGTGAGGAG TTTGAGGCTA TAGTGCAATG TGATCTTGCC TGTGAATAGC CACTGCACTC    1686
GAGCCTAGGC AACAAAGTGA GACCCCGTCT CCAAAAAAAA AAATGGTATA TTGGTATTTC    1746
TGTATATGAA CAATGAATGA TCTGAAAACA AGAAAATTCC ATTCACGATG GTATTAAAAA    1806
AATAAAATAC AAATAAAATT AGCAAAATAA TTATAAAACT TGTACATCGA AAATTTCAAA    1866
```

```
GCACTCTGAG GGAAATTAAA GATGATCTAA ATAATTGGAG AGACACTCTA TGATCACTGA 1925
TTGGAAAATT CATTCAATAT TGTTAAGATA ACAATTGTCC CCAAATTGAT GCATGCATTC 1985
AATTTAGTCT TCATCAAAAT TCCAGCCAGG TTTTTGCAGA AATTGACAAG CTGTACCCAA 2045
AATGTATATG GAAATGAAAA GACCCAGAAG AGCAAATAAT TTTTTAAAAA CAAAGTTGGA 2105
AAACTTTTAC TTCCTAATTT TAAAACTTAC TATAAACCTA AAGTTATCAA GACCATTTAG 2165
T                                                                 2167
```

(2) INFORMATION FOR SEQ ID NO: 8:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 1334 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: Genomic DNA

      (vi) ORIGINAL SOURCE:
          (A) ORGANISM: human
          (F) TISSUE TYPE: placenta

      (iX) FEATURE:
          (A) NAME/KEY: intron
          (B) LOCATION: 1..1334
          (C) IDENTIFICATION METHODS: E

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
GTATTTTTTT TAATTCGCAA ACATAGAAAT GACTAGCTAC TTCTTCCCAT TCTGTTTTAC   60
TGCTTACATT GTTCCGTGCT AGTCCCAATC CTCAGATGAA AAGTCACAGG AGTGACAATA  120
ATTTCACTTA CAGGAAACTT TATAAGGCAT CCACGTTTTT TAGTTGGGGT AAAAAATTGG  180
ATACAATAAG ACATTGCTAG GGGTCATGCC TCTCTGAGCC TGCCTTTGAA TCACCAATCC  240
CTTTATTGTG ATTGCATTAA CTGTTAAAA CCTCTATAGT TGGATGCTTA ATCCCTGCTT  300
GTTACAGCTG AAAATGCTGA TAGTTTACCA GGTGTGGTGG CATCTATCTG TAATCCTAGC  360
TACTTGGGAG GCTCAAGCAG GAGGATTGCT TGAGGCCAGG ACTTTGAGGC TGTAGTACAC  420
TGTGATCGTA CCTGTGAATA GCCACTGCAC TCCAGCCTGG GTGATATACA GACCTTGTCT  480
CTAAAATTAA AAAAAAAAA AAAAAAAACC TTAGGAAAGG AAATTGATCA AGTCTACTGT  540
GCCTTCCAAA ACATGAATTC CAAATATCAA AGTTAGGCTG AGTTGAAGCA GTGAATGTGC  600
ATTCTTTAAA AATACTGAAT ACTTACCTTA ACATATATTT TAAATATTTT ATTTAGCATT  660
TAAAGTTAA AAACAATCTT TTAGAATTCA TATCTTTAAA ATACTCAAAA AAGTTGCAGC  720
GTGTGTGTTG TAATACACAT TAAACTGTGG GGTTGTTTGT TTGTTTGAGA TGCAGTTTCA  780
CTCTGTCACC CAGGCTGAAG TGCAGTGCAG TGCAGTGGTG TGATCTCGGC TCACTACAAC  840
CTCCACCTCC CACGTTCAAG CGATTCTCAT GCCTCAGTCT CCCGAGTAGG TGGGATTACA  900
GGCATGCACC ACTTACACCC GGCTAATTTT TGTATTTTTA GTAGAGCTGG GGTTTCACCA  960
TGTTGGCCAG GCTGGTCTCA AACCCCTAAC CTCAAGTGAT CTGCCTGCCT CAGCCTCCCA 1020
AACAAACAAA CAACCCCACA GTTTAATATG TGTTACAACA CACATGCTGC AACTTTTATG 1080
AGTATTTTAA TGATATAGAT TATAAAAGGT TGTTTTTAAC TTTTAAATGC TGGGATTACA 1140
GGCATGAGCC ACTGTGCCAG GCCTGAACTG TGTTTTTAAA AATGTCTGAC CAGCTGTACA 1200
TAGTCTCCTG CAGACTGGCC AAGTCTCAAA GTGGGAACAG GTGTATTAAG GACTATCCTT 1260
TGGTTAAATT TCCGCAAATG TTCCTGTGCA AGAATTCTTC TAACTAGAGT TCTCATTTAT 1320
TATATTTATT TCAG                                                  1334
```

(2) INFORMATION FOR SEQ ID NO: 9:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 4773 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: Genomic DNA

      (vi) ORIGINAL SOURCE:
          (A) ORGANISM: human
          (F) TISSUE TYPE: placenta

      (iX) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 1..4773
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
GTAAGACTGA GCCTTACTTT GTTTTCAATC ATGTTAATAT AATCAATATA ATTAGAAATA    60
TAACATTATT TCTAATGTTA ATATAAGTAA TGTAATTAGA AAACTCAAAT ATCCTCAGAC   120
CAACCTTTTG TCTAGAACAG AAATAACAAG AAGCAGAGAA CCATTAAAGT GAATACTTAC   180
TAAAAATTAT CAAACTCTTT ACCTATTGTG ATAATGATGG TTTTTCTGAG CCTGTCACAG   240
GGGAAGAGGA GATACAACAC TTGTTTTATG ACCTGCATCT CCTGAACAAT CAGTCTTTAT   300
ATAAATAATA ATGTAGAATA CATATGTGAG TTATACATTT AAGAATAACA TGTGACTTTC   360
CAGAATGAGT TCTGCTATGA AGAATGAAGC TAATTATCCT TCTATATTTC TACACCTTTG   420
TAAATTATGA TAATATTTTA ATCCTAGTT GTTTTGTTGC TGATCCTTAG CCTAAGTCTT   480
AGACACAAGC TTCAGCTTCC AGTTGATGTA TGTTATTTTT AATGTTAATC TAATTGAATA   540
AAAGTTATGA GATCAGCTGT AAAAGTAATG CTATAATTAT CTTCAAGCCA GGTATAAAGT   600
ATTTCTGGCC TCTACTTTTT CTCTATTATT CTCCATTATT ATTCTCTATT ATTTTTCTCT   660
ATTTCCTCCA TTATTGTTAG ATAAACCACA ATTAACTATA GCTACAGACT GAGCCAGTAA   720
GAGTAGCCAG GGATGCTTAC AAATTGGCAA TGCTTCAGAG GAGAATTCCA TGTCATGAAG   780
ACTCTTTTTG AGTGGAGATT TGCCAATAAA TATCCGCTTT CATGCCCACC CAGTCCCAC    840
TGAAAGACAG TTAGGATATG ACCTTACTGA AGGTACCAAG GGGCAACTTG GTAGGGAGAA   900
AAAAGCCACT CTAAATATA ATCCAAGTAA GAACAGTGCA TATGCAACAG ATACAGCCCC   960
CAGACAAATC CCTCAGCTAT CTCCCTCCAA CCAGAGTGCC ACCCCTTCAG GTGACAATTT  1020
GGAGTCCCCA TTCTAGACCT GACAGGCAGC TTAGTTATCA AAATAGCATA AGAGGCCTGG  1080
GATGGAAGGG TAGGGTGGAA AGGGTTAAGC ATGCTGTTAC TGAACAACAT AATTAGAAGG  1140
GAAGGAGATG GCCAAGCTCA AGCTATGTGG GATAGAGGAA AACTCAGCTG CAGAGGCAGA  1200
TTCAGAAACT GGGATAAGTC CGAACCTACA GGTGGATTCT TGTTGAGGGA GACTGGTGAA  1260
AATGTTAAGA AGATGGAAAT AATGCTTGGC ACTTAGTAGG AACTGGGCAA ATCCATATTT  1320
GGGGGATCCT GAAGTTTATT CAATTTTCAT GGCCCTTTTA AATAAAAAGA ATGTGGCTGG  1380
GCGTGGTGGC TCACACCTGT AATCCCAGCA CTTTGGGAGG CCGAGGGGGG CGGATCACCT  1440
GAAGTCAGGA GTTCAAGACC AGCCTGACCA ACATGGAGAA ACCCCATCTC TACTAAAAAT  1500
ACAAAATTAG CTGGGCGTGG TGGCATATGC CTGTAATCCC AGCTACTCGG GAGGCTGAGG  1560
CAGGAGAATC TTTTGAACCC GGGAGGCAGA GGTTGCGATG AGCCTAGATC GTGCCATTGC  1620
ACTCCAGCCT GGGCAACAAG AGCAAAACTC GGTCTCAAAA AAAAAAAAAA AAAAGTGAAA  1680
TTAACCAAAG GCATTAGCTT AATAATTTAA TACTGTTTTT AAGTAGGGCG GGGGGTGGCT  1740
GGAAGAGATC TGTGTAAATG AGGGAATCTG ACATTTAAGC TTCATCAGCA TCATAGCAAA  1800
TCTGCTTCTG GAAGGAACTC AATAAATATT AGTTGGAGGG GGGAGAGAG TGAGGGGTGG  1860
ACTAGGACCA GTTTTAGCCC TTGTCTTTAA TCCCTTTTCC TGGCACTTAA AAGGATCTTA  1920
GCAGTGGTTA TAAAAGTGGC CTAGGTTCTA GATAATAAGA TACAACAGGC CAGGCACAGT  1980
GGCTCATGCC TATAATCCCA GCACTTTGGG AGGGCAAGGC GAGTGTCTCA CTTGAGATCA  2040
GGAGTTCAAG ACCAGCCTGG CCAGCATGGC GATATTCTGT CTCTACTAAA AAAAATACAA  2100
AAATTAGCCA GGCATGGTGG CATGCACCTG TAATCCCAGC TACTCGTGAG CCTGAGGCAG  2160
AAGAATCGCT TGAAACCAGG AGGTGTAGGC TGCAGTGAGC TGAGATCGCA CCACTGCACT  2220
CCAGCCTGGG CGACAGAATG AGACTTTGTC TCAAAAAAAG AAAAGATAC AACAGGCTAC  2280
CCTTATGTGC TCACCTTTCA CTGTTGATTA CTAGCTATAA AGTCTATAA AGTTCTTTGG  2340
TCAAGAACCT TGACAACACT AAGAGGGATT TGCTTTGAGA GGTTACTGTC AGAGTCTGTT  2400
TCATATATAT ACATATACAT GTATATATGT ATCTATATCC AGGCTTGGCC AGGGTTCCCT  2460
CAGACTTTCC AGTGCACTTG GGAGATGTTA GGTCAATATC AACTTTCCCT GGATTCAGAT  2520
TCAACCCTTT CTGATGTAAA AAAAAAAAAA AAAAGAAAG AAATCCCTTT CCCCTTGGAG  2580
CACTCAAGTT TCACCAGGTG GGGCTTTCCA AGTTGGGGGT TCTCCAAAGT CATTGGGATT  2640
GCTTTCATAT CCATTGCTA TGTACCTTCC CTATCATATT TGGGAGTGGT CAACATCAAA  2700
ATTAGGAAAG CTACTGCCCA AGGATGTGCT TACCTCTATT CTGAAATTTG CAATAAGGT  2760
GATTAAAGAG ATTGCCTGTT CTACCTATCC ACAATCCTGG TTTCAACTGT AACTTTCTTT  2820
TTTCTTTTTT TTCTTTTTTT CTTTTTTTTT CAAATGGAGT CTCGCTCTGT CGCCCAGGCT  2880
AGAGTGCAGT GGCACGATCT CAGCTCACTG CAAGCTCCGC CTCCCGGGTT CACGCCATTC  2940
TCCTGCCTCA CCCTCCCAAG CAGCTGGGAC TATAGGCGCC TGCCACCATG CCCAGTTAAT  3000
TTTTTGTATT TTTAGTAGAG ACGGGGTTTC ACCGTGTTAG CCAGGATGGT CTCGATCTCC  3060
TGAACTTGTG ATCCGCCCGC CTCAGCCTCC CAAAGTGCTG GGATTACAGG CGTGAGCCAT  3120
CGCACCCGGC TCAACTGTAA CTTTCTATAC TGGTTCATCT TCCCTGTAA TGTTACTAGA  3180
GCTTTTGAAG TTTTTGGCTAT GGATTATTTC TCATTTATAC ATTAGATTC AGATTAGTTC  3240
CAAATTGATG CCCACAGCTT AGGTCTCTT CCTAAAATTGT ATATTGTACA CAGCTGCAGA  3300
AGTGGGTGCC AATAGGGGAA CTAGTTTATA CTTTCATCAA CTTAGGACCC ACACTTGTTG  3350
ATAAAGAACA AAGGTCAAGA GTTATGACTA CTGATTCCAC AACTGATTGA GAAGTTGGAG  3420
ATAACCCCGT GACCTCTGCC ATCCAGAGTC TTTCAGGTAT CTTTGAAGGA TGAAGAAATG  3480
CTATTTTAAT TTTGGAGGTT TCTCTATCAG TGCTTAGGAT CATGGGAATC TGTGCTGCCA  3540
TGAGGCCAAA ATTAAGTCCA AAACATCTAC TGGTTCCAGG ATTAACATGG AAGAACCTTA  3600
GGTGGTGCCC ACATGTTCTG ATCCATCCTG CAAAATAGAC ATGCTGCACT AACAGGAAAA  3660
```

                                  - 36 -

```
GTGCAGGCAG CACTACCAGT TGGATAACCT GCAAGATTAT AGTTTCAAGT AATCTAACCA 3720
TTTCTCACAA GGCCCTATTC TGTGACTGAA ACATACAAGA ATCTGCATTT GGCCTTCTAA 3780
GGCAGGGCCC AGCCAAGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAACTG 3840
GAGTGCTTGG CAGGGAAGAC AGAGTCAAGG ACTGCCAACT GAGCCAATAC AGCAGGCTTA 3900
CACAGGAACC CAGGGCCTAG CCCTACAACA ATTATTGGGT CTATTCACTG TAAGTTTTAA 3960
TTTCAGGCTC CACTGAAAGA GTAAGCTAAG ATTCCTGGCA CTTTCTGTCT CTCTCACAGT 4020
TGGCTCAGAA ATGAGAACTG GTCAGGCCAG GCATGGTGGC TTACACCTGG AATCCCAGCA 4080
CTTTGGGAGG CCGAAGTGGG AGGGTCACTT GAGGCCAGGA GTTCAGGACC AGCTTAGGCA 4140
ACAAAGTGAG ATACCCCCTG ACCCCTTCTC TACAAAAATA AATTTTAAAA ATTAGCCAAA 4200
TGTGGTGGTG TATACTTACA GTCCCAGCTA CTCAGGAGGC TGAGGCAGGG GGATTGCTTG 4260
AGCCCAGGAA TTCAAGGCTG CAGTGAGCTA TGATTCACC ACTGCACTTC TGGCTGGGCA 4320
ACAGAGCGAG ACCCTGTCTC AAAGCAAAAA GAAAAGAAA CTAGAACTAG CCTAAGTTTG 4380
TGGAGGAGG TCATCATCGT CTTTAGCCGT GAATGGTTAT TATAGAGGAC AGAAATTGAC 4440
ATTAGCCCAA AAAGCTTGTG GTCTTTGCTG GAACTCTACT TAATCTTGAC CAAATGTGGA 4500
CACCACTCAA TGGGAGAGGA GAGAAGTAAG CTGTTTGATG TATAGGGGAA AACTAGAGGC 4560
CTGGAACTGA ATATGCATCC CATGACAGGG AGAATAGGAG ATTCGGAGTT AAGAAGGAGA 4620
GGAGGTCAGT ACTGCTGTTC AGAGATTTTT TTTATGTAAC TCTTGAGAAG CAAAACTACT 4680
TTTGTTCTGT TTGGTAATAT ACTTCAAAAC AAACTTCATA TATTCAAATT GTTCATGTCC 4740
TGAAATAATT AGGTAATGTT TTTTTCTCTA TAG                            4773
```

(2) INFORMATION FOR SEQ ID NO: 10:

      (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 8835 base pairs
         (B) TYPE: nucleic acid
         (C) STRANDEDNESS: double
         (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: Genomic DNA

      (vi) ORIGINAL SOURCE:
         (A) ORGANISM: human
         (F) TISSUE TYPE: placenta

      (ix) FEATURE:
         (A) NAME/KEY: intron
         (B) LOCATION: 1..8835
         (C) IDENTIFICATION METHODS: E

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
GTAAGAAATA TCATTCCTCT TTATTTGGAA AGTCAGCCAT GGCAATTAGA GGTAAATAAG   60
CTAGAAAGCA ATTGAGAGGA ATATAAACCA TCTAGCATCA CTACGATGAG CAGTCAGTAT  120
CAACATAAGA AATATAAGCA AAGTCAGAGT AGAATTTTTT TCTTTTATCA GATATGGGAG  180
AGTATCACTT TAGAGGAGAG GTTCTCAAAC TTTTTGCTCT CATGTTCCCT TTACACTAAG  240
CACATCACAT GTTAGCATAA GTAACATTTT TAATTAAAAA TAACTATGTA CTTTTTTAAC  300
AACAAAAAA AGCATAAAGA GTGACACTTT TTTATTTTTA CAAGTGTTTT AACTGGTTTA  360
ATAGAAGCCA TATAGATCTG CTGGATTCTC ATCTGCTTTG CATTCAGACT ACTGCAATAT  420
TGCACAGAAT GCAGCCTCTG GTAAACTCTG TTGTACACTC ATGAGAGAAT GGGTGAAAAA  480
GACAAATTAC GTCTTAGAAT TATTAGAAAT AGCTTTCACT TTAGGAACTC CCTGAGAATT  540
GCTGCTTTAG AGTGGTAAGA TAAATAAGCT TCTCTTTAAA CGGAATCTCA AGACAGAATC  600
AGTTACAGTA AAAGCAAACA AAAAATTTGC CCATGGTTAG TCATCTTGTG AAATCTGCCA  660
CACCTTTGGA CTGGGCTACA ATTGGATAAT ATAGCATTCC CCGAGATAAT TTTCTCTCAC  720
AATTAAGGAA AGGGCTGAAT AAATATCTCT GTTTGAAGTT GAATAACAAA AATTAGGACC  780
CCCTAAATTT TAGGGCTCCT GAAATTCGTC TTTTTGCCTA TATTCAGCTA CTTTACGTTC  840
TATTAAATCT TCTTTCAGGC CAGGTGCACT AGCTCATGCC TAGAATCTCA GGCAGGCCTG  900
AGCCCAGGAA TTTGAGACCA GCCAGGGCAA CACAGTCTCT ACAAAAAAAT AAAAAATTAC  960
CTGGGTGTGT TGGTGCATGC CTGTAGAACT ACTCAGGATG CTGAGGACTG CTTGAGCCCA 1020
GGATAGCCAA ATCTGTGGTG AGTTCAGCCA CTAAACAGAG CGAGACTTTC TCAAAAAAAC 1080
AAACAAAAAA ACAAACAAAC TTCCTTCAAA ATAACTTTTT ATCTGCAATG TTTTCCTATT 1140
GCCTGTGAGA TTAAATTTAC TCTTTTACCT GATTTCCAAA GCCCTCCATA ATCTAATCCG 1200
ACTTTACCTT GTGTTCCTG CAAAATAGCA GGACTGTTCC ACTACAATCC AAAAATCACA 1260
GGTTGGGTGC AGTGGCTCAC TCCTGTAATC CCAACACTTT GGAAGGCCAA GGCAGGTGGA 1320
TTGCTTCAGC TCAGGAGTTC AAGACCAGCC TGGGCAACAT GGCAAAAACC CTGTCTCTCC 1380
AAAACATACA AAAATTAGCC AGATGTGGTA GTATGTGCCT GTAGTCCCAA CTACTCAAAA 1440
GGCTAAGGCA AGAGGATCAC TTGAGCCCAG GAGGTCAAGG CTACAGTGAG CCATGTTTAC 1500
TGTGTCACTG CACTCCAGCC TGGGTGATAG AGCAAGACCA TGTCTCAAAA AAAAAAAAAA 1560
```

```
GAAAAGAAAA GAAAAAAACA TTGGTCTATT CAGTTCACCC CCACCACAAC ATTGTTTTGA 1620
TTATCACATA AATGCTGGTC CATTGCCTTC TCTATCTATT CAAATCTTTA AGCATTCTTT 1680
GAGATTCAAC TCAATTCTCC TTTTCAAACT AGGCCATTTA AACTACATCA GTTCCATTTT 1740
GATTTTCTTG CTTTGAGTCT ATAGACTCAA AAACAAAAAC TTAAAAACTT ATTTTTTAAG 1800
TTTTCTGCTA CTCTCACTTC TTCAACACTC ACATACACGC ATTCATAATA AGATGGCAGA 1860
ATGTTCAAGG ATAAAATGAT TTATAGAACT GAAAGTTAG GTTTTGATCT TGTTGCTGTC 1920
AAGATGACTA CCTACCTGAT CTCAGGTAAT TAATTATGTA GCATGCTCCC TCATTTCATC 1980
CCATACCTAT TCAACAGGAT TGGAATTCCA CAGCAAGGAT AAACATAATC ATAGTTGCTT 2040
TTCAAGTTCA AGGCATTTTA ACTTTTAATC TAGTAGTATG TTTGTTGTTG TTGTTGTTGT 2100
TTGAGATGGA GCCCTGCTGT GTCACCCAGG CTGGAGTGCA GTGGCACGAA CTCGGCTCAC 2160
TGCAACCTCT GCCTCATGGG TTCAATCAGT TATTCTGCCT CAGTGTCCCA AGTAGCTGGG 2220
ACTACAAGGC ACATGCCACC ATGCCTGGCT AATTTTTGTA TTTTAGTAG AAACAGGGCT 2280
TCACCATGTT GGCCAGGCTG GTCTCGAACT CCTGACCTCA AGTGATCCGC CCGCCTCGGC 2340
CTCCCAAAGT GCTGGGATTA CAGGCATAAG CCACCGTGCC CAGCCTAATA GTATGTTTTT 2400
AAACTCTTAG TGGCTTAACA ATGCTGGTTG TATAATAAAT ATGCCATAAA TATTTACTGT 2460
CTTAGAATTA TGAAGAAGTG GTTACTAGGC CGTTTGCCAC ATATCAATGG TTCTCTCCTT 2520
ACAGCTTTAA TTAGAGTCTA GAATTGCAGG TTGGTAGAGC TGGAACAGAC CTTAAAGATT 2580
GACTAGCCAA CTTCCTTGTC CAAATGAGGG AACTGAGACC CTTAAAAATTA AGTGACTTGC 2640
CCCAGACAAA ACTGGAACTC ATGTGTCCTA ATTCCATCA TGAATTCTA CCATTCACTA 2700
GCCTCTGCT AGTTGTCAAA GTATTGCATA ACTAAATTTT TATGTCTGTT TTAAAGAACA 2760
AATTGTCACT GCTTACTCCT GGGAGGGTCT TTCTGAGGTG GTTATATAACT CTTAAAAAAA 2820
AAAAAGTCAG TAGTCTGAGA ATTTTAGACG AAATAGTCAA AGCATTTTA TCCAATGGAT 2880
CTATAATTTT CATAGATTAG AGTTAAATCA AAGAAACACG GATGAGAAAG GAAGAGGAAA 2940
ATTGAGGAGA GGAGGAATGG GGATGAGAAC ACACTACTTG TAATCAGTCA TAGATGTACT 3000
GAGAACTAAC AAGAAGAATT GTAAGAAAAT AAGAATGAAG AATTCAAAT CAACACATGA 3060
AATAAAAAGA AACTACTAGG GAAAAATGGA GAAGACATTA GAAAAATTAT TCTATTTTTA 3120
AAATTCTGTT TTCAGGCTTC CCTCCTGTTC TTCCTCCTTC TCATTGGTTT TCAGGTGGAG 3180
GGAAAGTTTA AGATGGAAAA AATATATATA TTCTACACAT CCCTTTCTAC GCTGTTGTCA 3240
TGGCAACAAG GTTTATCATA GCAAACTTTT ATTCATACAA CATTTATTGA GTTCTTACTG 3300
TGTGGTAAGC TCTTTCCAGG TGTTGAAAAT TCAGGGGAAA AAGACAACT CATTGTCTTA 3360
AAACTCAGAT GAAAGCTGAA CAGACCTATT TTTAATCAAA GTAATCTCAA TTTAGGGTAG 3420
TAAGAGCTAT TTAAGAAGCA TGAACAGGTG TGAAGGAGGT AGGACTCTGA GGAGAGAATA 3480
GTTAGCTAGG AATGAAAGAG CAGAGAAGTT TTCCTAGAGG AACTATTAAA GCTGGGAGTT 3540
ACGGGATGAA AGATGAGGCA GGGTTTGCAG GCAAAAAAAA AAAAAAGGCA GGGGAAGGGG 3600
AAGTTCTGGC CTGGCAGAGA GAATAACTGT GGCAACAATG GAGGAGAGTC TGGAAGCAAG 3660
AAAACCAAGT AGAAGAGTAT TAAAATAGAA GATGCCAGGG GTAATGAGGG CTTGATTTAA 3720
AACAGTGCTG TTGGAGATGG AGAGGAGATA CCAAATTCTG GAGACATTTC TGAGTTAGAA 3780
CCTACAGTAT TTATCAGACA AGGGAAAGAT TAGACAAAG3 AGTTAAGAAT GACTCCCAGG 3840
TTTCAGTTTG GGGCAGGTAA CTAGGACATG TTTTGAAAAG TAATGTATTG GATCTCTTAC 3900
CATTGGAACT ATGTATGTGG AGCCAAATTA AAATTTGTAC ATGTATATAA CTCTCCCCCC 3950
ACCACCAGTA ACTACTTCCC TAACTCTCTA CTTTGTAGCC AGACTTCCTA AAAGAATAGT 4020
TTGTAGTCAC TGTCTTTACT TTTCCCCTCC CATTCTGTCC TAGATATTTG TCCACCTACC 4080
ATCTGCTGCC TCCACTTTAC CCAAACTGTT CTACGGTTGC CCAAAACTTC CTAATTGCCA 4140
AATTCAATGA ACAAGTTTAA GCTTATATGT AAATTAGGAG CTCTACAGTT TGATTTCGAG 4200
CAGCCCCTCC TGAAACCCTT TCTTTTTCGA CTTCTGTGAC ACATCTCAGA TTTACAAAAC 4260
TGAACTAATT ATTTTACACT TGAGCTGTAT TTTGGTTCTT CTTTCTTGAT GAATGAGGTA 4320
ACCACTCAAC AAAATTGCCCA AGCCAAAAAC TACGAAGTCA TCCTCAGTTC CTCCTTCTTC 4380
TGTTTGACCC ACAACAGATC AGCTGAGAAA TCCCGCTGTT TAGTATCTCT TGAATTCATT 4440
ACCTTAATTT ATAGCCTCAT CAATTCTTAA TTGTTAAAAT TACTTCAGTA GTTGTTGTCT 4500
GACCTCTGTC CAATCTTGTT CAATCAGGTC CATTCTTTTG TTCTTGGTGG TGGTGGTGGT 4560
GTTGATAGAG TTTGGCTTTT GCTGCCCAGG CTGAAGTGCA GTGGAGCACT TCACTGCAAC 4620
CACAGCCTCC TGGGTTTAAG CAGTTCACCC TCCGAGTAG CTGGGACTAC AGGTATGTGC 4680
CACCACACCC AGCTAATTTT GTGTTTTCAG TAGAGACAGG GTTTCACCAT GTTGGTCAGG 4740
CTGGTCTCAA ACTCCTGACC TCAAGCAATC CACCCACCTC AGCCTCCCAA AGTGTTGGGA 4800
TTACAGGCAT GAGCCACTGC ACAGGAGCCA GATCCATTGT TTATGTTGCT TCTAGAGTGA 4860
GTTTTTAAAA CACAAATTTG ACCATATCTT TCTCCAATTT AAGTCAGTAT TTTTTTTTTC 4920
AGGAAAAAAC AGTTCAAACT CTTTAGTCGG CTTACACAAG GCCTTTGTAG TCTGACTCTT 4980
CTTTCCAAGC TTTCATCAAA GTATACTGCA AGTTACATTT TATGTGAATT GAATTAGGCA 5040
ACGGTATAAA AATTATAGTT TATATGGGCA AAATGGAAAT AATGTTAACT CTTCCCAAATA 5100
GTTTATCTAG AATGACATAA TTTCAAAGCT GTCAGTCAA ATGAGTTATA AACTGTTAAC 5160
ACTATTGCCA CATGCAAGTG TCTCTTATAC TTGGTAGAAT TATCTGTTC CATGTCATTA 5220
TTATGTAAAT TAGACTTTAA ATAACTACAGA AGTTCTTCAG ACATACAGGT TATTATTGTG 5280
CTTTTTAAAC ATAATTTTAA ATAATTTTAT ATATGATAAT GTTATCCAAG TGCTAAGGGA 5340
TGTATTGTTA CTGCTGTGCA AAAAAAAAAA AAAAAAAAAC TCCAAATAAA TATGTTGAAA 5400
CCAAGTTTAT ATGCAAGAAA ACAATATTAA AAAGGCCAAA GTACCACCAT AATAGGCTGT 5460
GTGGAGACGG CAGGCTACAA AACACTAGTA ATAATGCTGA GAAAGTTGAA AAAAGAAAGA 5520
AAGCAACAAT ATGCTTTGGT TGTTGTAGGT TTATGTACTC CAAGAATATC TCCTCTCAAA 5580
```

```
CTTTTATGTT TTTTCCAAAG AAAAGTTAAC TTTGGCTGGG CGCAGTGGCT CTTGCCTGTA 5640
GTCCCAGCCT TTGGGAGGCC AAGGCGGGCA GATCACCTGA GGTCAGGAGT TTGAGACCAG 5700
CCTGACCAAA AATGGAGAAA CCCGCCCCCC TCACTACTAA AAGAATACAA AATTAGGCCG 5760
GGCACAGTGG CTTACCCCTG TCATCCCAGC ACTTTGGGAG GCCGAAGCAG GAAGATCACC 5820
TGAGGTCAGG AGTTCGAGAC CAGCCATGGA GAAACCCGTC TCTACTAAAA ATACAAAATT 5880
AGCCGGGCGT GGTGGTGCAT GACTGTAATC CCAGCTACTC AGGAGGCTAA GGCAGAGAAT 5940
CACTTGAACC CAGGCAGTGG AGGTTGCAGT GAGCCGAGAT CGTGCCATTG CACTCCAGCC 6000
TGGGCAACAA GAGCGAAACT CTGTATCCAA AAAACAAAAG AAAAGAAAAG GTAACCTTGA 6060
ACTATGTGAG ATCTTTAGAA ATGCATTCTT TCTGTAAAAT GTGACTACAT TTGCCTTATT 6120
TATGGTAAAA ATGTTGAGGC CTCAAACAAC CCATATTTTC TCGGTCTCCC CGCTGCCTAG 6180
CCTTTGTTCA CATTGCTTCT TCTTGGTGGA AGCTCTTCCT CTGGCCTTGA AAATGCCTGC 6240
TTCTCTTTTA AGGTAGCACA GTCATCACTT TCTGTGGTAA CCTTCTCCAG CACCATCAAA 6300
CAGAAAGAAT GAATCTCTTG TAAATTCAGC TCTTACGTCA TTCATTACAT TATTTTGTAA 6360
CTCTTTATAG ATTCTTCTCT CCCACTAGAC TCTGAGTCAC TGGAGAGTAG GAGCCAACTC 6420
TCATTCATGT GTGGTTTGGT CAGCTACTGG CCACATTCCT GATGCATAGT TAATGCTCAA 6480
ACCTTAACTG GTGAATCAGC TCAAATATTG TCCTTCTCTA AATCCATTCA CTCATTGACT 6540
AACTATGTAC TCAAAATAGT AAACACCAGT AATTTAATCC AATTCCTGCC CATACTGCTT 6600
GGTACATTTC AGGTGAATTA GTTTGATAAA TATGTGTGTA TTACATAATA TTAAAGTATG 6660
TACAGAAGAT CATGCTAATC ATAATTCACA ACTGATAACT AATCAAACAT AAATGCTCTC 6720
AGGTTAACAA ATGTCTGCCT TCTCAGTTAA TGCAGTCATT AACAAACACC TTCTGATGCT 6780
GATAATAGGG CCTTGTTCAG CAATGAAGCC ATAAAGGTGA ATAAAGAACA TGCCCTCGTG 6840
GAGCTCACAG CCTAGTCATT ATTGTTCTGA TTTTTAATAT TAATGTTGGT TTGGGTTTTG 6900
GTGAAAAATG TTTAGACTTA TCTTAGTGAT CTTTTCATCC TTTGCTATAT TATTTTTCTC 6960
TAAGAGTCTT CCTTATCCCC TCCTTTAAAA AACTAGGTGA TAATTCTAAA TTGTAAATTT 7020
AAATATTATA AATAGCTTAT AAAATTTAAT ATTTATAATA TTTAAATGTT TGATAAATAT 7080
TTAAATTTTA TAATATTTAA ATGTTTATTT AAATTCATTT GTACATCAGT TTTTATTTTA 7140
TTTAAATGTG TTGGCCAGGC ATGGTGGCTG ACACCTATAA TCCCAGAACT TTGAGAGGCC 7200
AAGTCAGGCA AACCATTTGA GCTCAGGAGT TTGAGACCAC CCTGGGCAAC GTGGTGAAAC 7260
CCTGTCTCTA CCAAACATAT GAAAACTTAT CTGGGTGTGG TGGCACGCAT CTGTGGTCCC 7320
AGATGGGAGT CCCAGGCTAA GATGGGAGAA TCGCTTGAAC CCAGGTGAGA GGGGTGGGGT 7380
GGATGTTGCA GTGAGCTGAG ATCGTGCCAC TGCACTCCAA CCTGGGTGAC AGAGTGAGAC 7440
TCCATCTCAA AAAAAAAAAA TGTTATCTAA ATAAGATAAA TTTAATAACT GTTCGCACTT 7500
AGATGAGCAT AAGGAACTAA ACCTAGATAA AACTATCAAA TAAGGCCTGG GTACAGTGAC 7560
TCATGCCTGT AATCTCAAGC ACTTTGGGAG GCCAAAATTA TACAAAGTTA GTTGTATAAC 7620
ACCAACTAAC AACTATTTTG GGGTTAGCTT AATTCAGATT AATTTTTTTT AAACTGAGTT 7680
TTAAATTCCT GCTTACTCTA CCATACATGC TAGGCCTCAT ATTATGCTAG AAAAATTTTG 7740
AGCACAGATT TATGAATACT CTCCTGCATA CCATTTAATT TTTAAACAAA TTTTAATGCA 7800
GTATATATGT GCCTTTTTAC CAACACATTA AATAATAAGA TCTACTGTGA GGACTAAATT 7860
TCTGTAATTT CAAAGTAGTA ATGAGTTTAA ACCATGTCTC AAGATCTCTG CAATAACTGT 7920
AGCACAACAG AAAATAGGTA TTTCTATTAA TGACAGAGTC ACAAGTACTA CTAATAATAC 7980
TGTGGTTTGT TTCCTGCAAC TAATCATGGG AGGAATGCTA AATTTCAGAG GTTGGTGAAA 8040
ATACATGTGT ATTTTTTTCC TCATCCAAGT TCACAGATTT CTCACACTGA GAACTCCTAT 8100
TCCATAACAA AATTCTGGAA GGCTGCACAC CGTATTGGAA GAAGGGCAGA AAGGAAAAGC 8160
AAATGGAAGG ATTTAAATTT TTTTCAAATC CTGTATCCCT TGATTTTACA GCAAGATTGT 8220
ATTTATGTAT TACTTGTGTT AAAAATATAG TATAATCGAG ACTCCAGATC AAAAATCACC 8280
GTAGCTCAGG GAGAAAGAGG GCCACCAAAT GCCAGAGCCC TTCAGCCTTC TCCCACCCTG 8340
CCTGTACCCT CAGATGGAAG CACTTTTTTA TCATTGTTTC ACCTTTAGCA TTTTGACAAT 8400
GAAGTCACAA ACCTTCAGCC TCTCACCCAT AGGAACCCAC TGGTTGTAAG AGAAGGATGA 8460
AGCCAGTCCT TCCTAAAGGG CACGATTAGA TGTGTTTATG GCATCCTCAG GTGAAACTAT 8520
ATTTATATTG ACAATATATT TATATTTCTC AAGGAATACT AGAATAATGA TTCAGTTCAG 8580
TACTAGGTCA TTTATCTATC CTTTATAATA TTGTTTAATG AGAAAATGCT TTCTATCTTC 8640
CAAATATTTG ATGATTTGTA AGAGAATACT TAAACATGGG TATTCATAAG CTGAAACTTC 8700
TGGCATTTAT TGAATGTCAA GATTGTTCAT CAGTATACTA GGTGATTAAC TGACCACTGA 8760
ACTTGAAGGT AGTATAAAGT AGTAGTAAAA GGTACAATCA TTGTCTCTTA ACAGATGGCT 8820
CTTTGCTTTC ATTAG                                                  8835
```

(2) INFORMATION FOR SEQ ID NO:11:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 1371 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: double
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: Genomic DNA

        (vi) ORIGINAL SOURCE:

```
            (A)  ORGANISM: human
            (F)  TISSUE TYPE: placenta

      (iX)  FEATURE:
            (A)  NAME/KEY: intron
            (B)  LOCATION: 1..1371
            (C)  IDENTIFICATION METHODS: E

      (xi)  SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTAAGGCTAA TGCCATAGAA CAAATACCAG GTTCAGATAA ATCTATTCAA TTAGAAAAGA     60
TGTTGTGAGG TGAACTATTA AGTGACTCTT TGTGTCACCA AATTTCACTG TAATATTAAT    120
GGCTCTTAAA AAAATAGTGG ACCTCTAGAA ATTAACCACA ACATGTCCAA GGTCTCAGCA    180
CCTTGTCACA CCACGTGTCC TGGCACTTTA ATCAGCAGTA GCTCACTCTC CAGTTGGCAG    240
TAAGTGCACA TCATGAAAAT CCCAGTTTTC ATGGGAAAAT CCCAGTTTTC ATTGGATTTC    300
CATGGGAAAA ATCCCAGTAC AAAACTGGGT GCATTCAGGA AATACAATTT CCCAAAGCAA    360
ATTGGCAAAT TATGTAAGAG ATTCTCTAAA TTTAGAGTTC CGTGAATTAC ACCATTTTAT    420
GTAAATATGT TTGACAAGTA AAAATTGATT CTTTTTTTTT TTTTCTGTTG CCCAGGCTGG    480
AGTGCAGTGG CACAATCTCT GCTCACTGCA ACCTCCACCT CCTGGGTTCA AGCAATTCTC    540
CTGCCTCAGC CTTCTGAGTA GCTGGGACTA CAGGTGCATC CCGCCATGCC TGGCTAATTT    600
TTGGGTATTT TTACTAGAGA CAGGGTTTTG GCATGTTGTC CAGGCTGGTC TTGGACTCCT    660
GATCTCAGAT GATCCTCCTG GCTCGGGCTC CCAAAGTGCT GGGATTACAG GCATGAACCA    720
CCACACATGG CCTAAAAATT GATTCTTATG ATTAATCTCC TGTGAACAAT TTGGCTTCAT    780
TTGAAAGTTT GCCTTCATTT GAAACCTTCA TTTAAAAGCC TGAGCAACAA AGTGAGACCC    840
CATCTCTACA AAAAACTGCA AAATATCCTG TGGACACCTC CTACCTTCTG TGGAGGCTGA    900
AGCAGGAGGA TCACTTGAGC CTAGGAATTT GAGCCTGCAG TGAGCTATGA TCCCACCCCT    960
ACACTCCAGC CTGCATGACA GTAGACCCTG ACACACACAC ACAAAAAAAA ACCTTCATAA   1020
AAAATTATTA GTTGACTTTT CTTAGGTGAC TTTCCGTTTA AGCAATAAAT TTAAAAGTAA   1080
AATCTCTAAT TTTAGAAAAT TTATTTTTAG TTACATATTG AAATTTTTAA ACCCTAGGTT   1140
TAAGTTTTAT GTCTAAATTA CCTGAGAACA CACTAAGTCT GATAAGCTTC ATTTTATGGG   1200
CCTTTTGGAT GATTATATAA TATTCTGATG AAAGCCAAGA CAGACCCTTA AACCATAAAA   1260
ATAGGAGTTC GAGAAAGAGG AGTAGCAAAA GTAAAAGCTA GAATGAGATT GAATTCTGAG   1320
TCGAAATACA AAATTTTACA TATTCTGTTT CTCTCTTTTT CCCCCTCTTA G            1371


(2)  INFORMATION FOR SEQ ID NO: 12:

      (i)   SEQUENCE CHARACTERISTICS:
            (A)  LENGTH: 3383 base pairs
            (B)  TYPE: nucleic acid
            (C)  STRANDEDNESS: double
            (D)  TOPOLOGY: linear

      (ii)  MOLECULE TYPE: Genomic DNA

      (vi)  ORIGINAL SOURCE:
            (A)  ORGANISM: human
            (F)  TISSUE TYPE: placenta

      (iX)  FEATURE:
            (A)  NAME/KEY: intron
            (B)  LOCATION: 1..3383
            (C)  IDENTIFICATION METHODS: E

      (xi)  SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GTAAAGTAGA AATGAATTTA TTTTTCTTTG CAAACTAAGT ATCTGCTTGA GACACATCTA     60
TCTCACCATT GTCAGCTGAG GAAAAAAAAA AATGGTTCTC ATGCTACCAA TCTGCCTTCA    120
AAGAAATGTG GACTCAGTAG CACAGCTTTG GAATGAAGAT GATCATAAGA GATACAAAGA    180
AGAACCTCTA GCAAAGATGG CTTCTCTATG CCTTAAAAAA TTCTCCAGCT CTTAGAATCT    240
ACAAAATAGA CTTTGCCTGT TTCATTGGTC CTAAGATTAG CATGAAGCCA TGGATTCTGT    300
TGTAGGGGGA GCGTTGCATA GGAAAAAGGG ATTGAAGCAT TAGAATTGTC CAAAATCAGT    360
AACACCTCCT CTCAGAAATG CTTTGGGAAG AAGCCTGGAA GGTTCCGGGT TGGTGGTGGG    420
GTGGGGCAGA AAATTCTGGA AGTAGAGGAG ATAGGAAGTG GTGGGGCAAG AAGACCACAT    480
TCAGAGGCCA AAAGCTGAAA GAAACCATGG CATTTATGAT GAATTCAGGG TAATTCAGAA    540
TGGAAGTAGA GTAGGAGTAG GAGACTGGTG AGAGGAGCTA GAGTGATAAA CAGGGTGTAG    600
AGCAAGACGT TCTCTCACCC CAAGATGTGA AATTTGGACT TTATCTTGGA GATAATAGGG    660
TTAATTAAGC ACAATATGTA TTAGCTAGGG TAAAGATTAG TTTGTTGTAA CAAAGACATC    720
```

```
CAAAGATACA GTAGCTGAAT AAGATAGAGA ATTTTTCTCT CAAAGAAAGT CTAAGTAGGC   780
AGCTCAGAAG TAGTATGGCT GGAAGCAACC TGATGATATT GGGACCCCCA ACCTTCTTCA   840
GTCTTGTACC CATCATCCCAG TAGTTGTTGA TCTCACTCAC ATAGTTGAAA ATCATCATAC   900
TTCCTGGGTT CATATCCCAG TTATCAAGAA AGGGTCAAGA GAAGTCAGGC TCATTCCTTT   960
CAAAGACTCT AATTGGAAGT TAAACATATC AATCCCCCTC ATATTCCATT GACTAGAATT  1020
TAATCACATG GCCACACCAA GTGCAAGGAA ATCTCGAAAA TATAATCTTT ATTCCAGGTA  1080
GCCATATGAC TCTTTAAAAT TCAGAAATAA TATATTTTTA AAATATCATT CTGGCTTTGG  1140
TATAAAGAAT TGATGGTGTG GGGTGAGGAG GCCAAAATTA AGGGTTGAGA GCCTATTATT  1200
TTAGTTATTA CAAGAAATGA TGGTGTCATG AATTAAGGTA GACATAGGGG AGTGCTGATG  1260
AGGAGCTGTG AATGGATTTT AGAAACACTT GAGAGAATCA ATAGGACATG ATTTAGGGTT  1320
GGATTTGGAA AGGAGAAGAA AGTAGAAAAG ATGATGCCTA CATTTTTCAC TTAGGCAATT  1380
TGTACCATTC AGTGAAATAG GGAACACAGG AGGAAGAGCA GGTTTTGGTG TATACAAAGA  1440
GGAGGATGGA TGACGCATTT CGTTTTGGAT CTGAGATGTC TGTGGAACGT CCTAGTGGAG  1500
ATGTCCACAA ACTCTTCTAC ATGTGGTTCT GAGTTCAGGA CACAGATTTG GGCTGGAGAT  1560
AGAGATATTG TAGGCTTATA CATAGAAATG GCATTTGAAT CTATAGAGAT AAAAAGACAC  1620
ATCAGAGGAA ATGTGTAAAG TGAGAGAGGA AAAGCCAAGT ACTGTGCTGG GGGGAATACC  1680
TACATTTAAA GGATGCAGTA GAAAGAAGCT AATAAACAAC AGAGAGCAGA CTAACCAAAA  1740
GGGGAGAAGA AAAACCAAGA GAATTCCACC GACTCCCAGG AGAGCATTTC AAGATTGAGG  1800
GGATAGGTGT TGTGTTGAAT TTTGCAGCCT TGAGAATCAA GGGCCAGAAC ACAGCTTTTA  1860
GATTTAGCAA CAAGGAGTTT GGTGATCTCA GTGAAAGCAG CTTGATGGTG AAATGGAGGC  1920
AGAGGCAGAT TGCAATGAGT GAAACAGTGA ATGGGAAGTG AAGAAATGAT ACAGATAATT  1980
CTTGCTAAAA GCTTGGCTGT TAAAAGGAGG AGAGAAACAA GACTAGCTGC AAAGTGAGAT  2040
TGGGTTGATG GAGCAGTTTT AAATCTCAAA ATAAAGAGCT TTGTGCTTTT TTGATTATGA  2100
AAATAATGTG TTAATTGTAA CTAATTGAGG CAATGAAAAA AGATAATAAT ATGAAAGATA  2160
AAAATATAAA AACCACCCAG AAATAATGAT AGCTACCATT TTGATACAAT ATTTCTACAC  2220
TCCTTTCTAT GTATATATAC AGACACAGAA ATGCTTATAT TTTTATTAAA AGGGATTGTA  2280
CTATACCTAA GCTGCTTTTT CTAGTTAGTG ATATATATGG ACATCTCTCC ATGGCAACGA  2340
GTAATTGCAG TTATATTAAG TTCATGATAA TTCACAATAA GGGCATATCT TTGCCCTTTT  2400
TATTTAATCA ATTCTTAATT GGTGAATGTT TGTTTCCAGT TTGTTGTTGT TATTAACAAT  2460
GTTCCCATAA GCATTCCTGT ACACCAATGT TCACACATTT GTCTGATTTT TTCTTCAGGA  2520
TAAAACCCAG GAGGTAGAAT TGCTGGGTTG ATAGAAGAGA AAGGATGATT GCCAAATTAA  2580
AGCTTCAGTA GAGGGTACAT GCCGAGCACA AATGGGATCA GCCCTAGATA CCAGAAATGG  2640
CACTTTCTCA TTTCCCCTTG GGACAAAAGG GAGAGAGGCA ATAACTGTGC TGCCAGAGTT  2700
AAATTTGTAC GTGGAGTAGC AGGAAATCAT TTGCTGAAAA TGAAAACAGA GATGATGTTG  2760
TAGAGGTCCT GAAGAGAGCA AAGAAAATTT GAAATTGCGG CTATCAGCTA TGGAAGAGAG  2820
TGCTGAACTG GAAAACAAAA GAAGTATTGA CAATTGGTAT GCTTGTAATG GCACCGATTT  2830
GAACGCTTGT GCCATTGTTC ACCAGCAGCA CTCAGCAGCC AAGTTTGGAG TTTTGTAGCA  2940
GAAAGACAAA TAAGTTAGGG ATTTAATATC CTGGCCAAAT GGTAGACAAA ATGAACTCTG  3000
AGATCCAGCT GCACAGGGAA GGAAGGGAAG ACGGGAAGAG GTTAGATAGG AAATACAAGA  3050
GTCAGGAGAC TGGAAGATGT TGTGATATTT AAGAACACAT AGAGTTGGAG TAAAAGTGTA  3120
AGAAAACTAG AAGGGTAAGA GACCGGTCAG AAAGTAGGCT ATTTGAAGTT AACACTTCAG  3180
AGGCAGAGTA GTTCTGAATG GTAACAAGAA ATTGAGTGTG CCTTTGAGAG TAGGTTAAAA  3240
AACAATAGGC AACTTTATTG TAGCTACTTC TGGAACAGAA GATTGTCATT AATAGTTTTA  3300
GAAAACTAAA ATATATAGCA TACTTATTTG TCAATTAACA AAGAAACTAT GTATTTTTAA  3360
ATGAGATTTA ATGTTTATTG TAG                                          3383
```

(2) INFORMATION FOR SEQ ID NO: 13:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 11464 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: Genomic DNA

      (vi) ORIGINAL SOURCE:
          (A) ORGANISM: human
          (F) TISSUE TYPE: placenta

      (ix) FEATURE:
          (A) NAME/KEY: 5'UTR
          (B) LOCATION: 1..3
          (C) IDENTIFICATION METHODS: E
          (A) NAME/KEY: leader peptide
          (B) LOCATION: 4..82
          (C) IDENTIFICATION METHODS: S

```
        (A)  NAME/KEY: intron
        (B)  LOCATION: 83..1453
        (C)  IDENTIFICATION METHODS: E
        (A)  NAME/KEY: leader peptide
        (B)  LOCATION: 1454..1465
        (C)  IDENTIFICATION METHODS: S
        (A)  NAME/KEY: intron
        (B)  LOCATION: 1466..4848
        (C)  IDENTIFICATION METHODS: E
        (A)  NAME/KEY: leader peptide
        (B)  LOCATION: 4849..4865
        (C)  IDENTIFICATION METHODS: S
        (A)  NAME/KEY: mat peptide
        (B)  LOCATION: 4866..4983
        (C)  IDENTIFICATION METHODS: S
        (A)  NAME/KEY: intron
        (B)  LOCATION: 4984..6317
        (C)  IDENTIFICATION METHODS: E
        (A)  NAME/KEY: mat peptide
        (B)  LOCATION: 6318..6451
        (C)  IDENTIFICATION METHODS: S
        (A)  NAME/KEY: intron
        (B)  LOCATION: 6452..11224
        (C)  IDENTIFICATION METHODS: E
        (A)  NAME/KEY: mat peptide
        (B)  LOCATION: 11225..11443
        (C)  IDENTIFICATION METHODS: S
        (A)  NAME/KEY: 3'UTR
        (B)  LOCATION: 11444..11464
        (C)  IDENTIFICATION METHODS: E


    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:


AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA        48
    Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala
    -35                 -30                 -25
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G   GTAAGG CTAATGCCAT      98
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
    -20                 -15                 -10
AGAACAAATA CCAGGTTCAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAACT     158
ATTAAGTGAC TCTTTGTGTC ACCAAATTTC ACTGTAATAT TAATGGCTCT TAAAAAAATA     218
GTGGACCTCT AGAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCACGT     278
GTCCTGGCAC TTTAATCAGT AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA     338
AAATCCCAGT TTTCATGGGA AAATCCCAGT TTTCATTGGA TTTCCATGGG AAAAATCCCA     398
GTACAAAACT GGGTCATTC AGGAAATACA ATTTCCCAAA GCAAATTGC AAATTATGTA       458
AGAGATTCTC TAAATTTAGA GTTCCGTGAA TTACACCATT TTATGTAAAT ATGTTTGACA     518
AGTAAAAATT GATTCTTTT TTTTTTTCT GTTGCCCAGG CTGGAGTGCA GTGGCACAAT       578
CTCTGCTCAC TGCAACCTCC ACCTCCTGGG TTCAAGCAAT TCTCCTGCCT CAGCCTTCTG     638
AGTAGCTGGG ACTACAGGTG CATCCCGCCA TGCCTGGCTA ATTTTTGGGT ATTTTTACTA     698
GAGACAGGGT TTTGCCATGT TGTCCAGGCT GGTCTTGGAC TCCTGATCTC AGATGATCCT     758
CCTGGCTCGG GCTCCCAAAG TGCTGGGATT ACAGGCATGA ACCACCACAC ATGGCCTAAA     818
AATTGATTCT TATGATTAAT CTCCTGTGAA CAATTTGGCT TCATTGAAA GTTTGCCTTC      878
ATTTGAAACC TTCATTTAAA AGCCTGAGCA ACAAAGTGAG ACCCCATCTC TACAAAAAAC     938
TGCAAAATAT CCTGTGGACA CCTCCTACCT TCTGTGGAGG CTGAAGCAGG AGGATCACTT     998
GAGCCTAGGA ATTTGAGCCT GCAGTGAGCT ATGATCCCAC CCCTACACTC CAGCCTGCAT    1058
GACAGTAGAC CCTGACACAC ACACACAAAA AAAAACCTTC ATAAAAAATT ATTAGTTGAC    1118
TTTTCTTAGG TGACTTTCCG TTTAAGCAAT AAATTTAAAA GTAAAATCTC TAATTTTAGA    1178
AAATTTATTT TTAGTTACAT ATTGAAATTT TTAAACCCTA GGTTTAAGTT TTATGTCTAA    1238
ATTACCTGAG AACACACTAA GTCTGATAAG CTTCATTTTA TGGGCCTTTT GGATGATTAT    1298
ATAATATTCT GATGAAAGCC AAGACAGACC CTTAAACCAT AAAAATAGGA GTTCGAGAAA    1358
GAGGAGTAGT AAAAGTAAAA GCTAGAATGA GATTGAATTC TGAGTCGAAA TACAAAATTT    1418
TACATATTCT GTTTCTCTCT TTTTCCCCCT CTTAG   CT GAA GAT GAT G   GTAAA    1470
                                         Ala Glu Asp Asp Glu
                                             -10
GTAGAAATGA ATTTATTTTT CTTTGCAAAC TAAGTATCTG CTTGAGACAC ATCTATCTCA    1530
CCATTGTCAG CTGAGGAAAA AAAAAAATGG TTCTCATGCT ACCAATCTGC CTTCAAAGAA    1590
ATGTGGACTC AGTAGCACAG CTTTGGAATG AAGATGATCA TAAGAGATAC AAAGAAGAAC    1650
```

- 42 -

```
CTCTAGCAAA AGATGCTTCT CTATGCCTTA AAAAATTCTC CAGCTCTTAG AATCTACAAA   1710
ATAGACTTTG CCTGTTTCAT TGGTCCTAAG ATTAGCATGA AGCCATGGAT TCTGTTGTAG   1770
GGGGAGCGTT GCATAGGAAA AAGGGATTGA AGCATTAGAA TTGTCCAAAA TCATCTAACAC  1830
CTCCTCTCAG AAATGCTTTG GGAAGAAGCC TGGAAGGTTC CGGGTTGGTG GTGGGGTGGG   1890
GCAGAAAATT CTGGAAGTAG AGGAGATAGG AATGGGTGGG GCAAGAAGAC CACATTCAGA   1950
GGCCAAAAGC TGAAAGAAAC CATGGCATTT ATGATGAATT CAGGGTAATT CAGAATGGAA   2010
GTAGAGTAGG AGTAGGAGAC TGGTGAGAGG AGCTAGAGTG ATAAACAGGG TGTAGAGCAA   2070
GACGTTCTCT CACCCCAAGA TGTGAAATTT GGACTTTATC TTGGAGATAA TAGGGTTAAT   2130
TAAGCACAAT ATGTATTAGC TAGGGTAAAG ATTAGTTTGT TGTAACAAAG ACATCCAAAG   2190
ATACAGTAGC TGAATAAGAT AGAGAATTTT TCTCTCAAAG AAAGTCTAAG TAGGCAGCTC   2250
AGAAGTAGTA TGGCTGGAAG CAACCTGATG ATATTGGGAC CCCCAACCTT CTTCAGTCTT   2310
GTACCCATCA TCCCCTAGTT GTTGATCTCA CTCACATAGT TGAAAATCAT CATACTTCCT   2370
GGGTTCATAT CCCAGTTATC AAGAAAGGGT CAAGAGAAGT CAGGCTCATT CCTTTCAAAG   2430
ACTCTAATTG GAAGTTAAAC ACATCAATCC CCTCATATT CCATTGACTA GAATTTAATC    2490
ACATGGCCAC ACCAAGTGCA AGGAAATCTG GAAAATATAA TCTTTATTCC AGGTAGCCAT   2550
ATGACTCTTT AAAATTCAGA AATAATATAT TTTTAAAATA TCATTCTGGC TTTGGTATAA   2610
AGAATTGATG GTGTGGGGTG AGGAGGCCAA AATTAAGGGT TGAGAGCCTA TTATTTTAGT   2670
TATTACAAGA AATGATGGTG TCATGAATTA AGGTAGACAT AGGGGAGTGC TGATGAGGAG   2730
CTGTGAATGG ATTTTAGAAA CACTTGAGAG AATCAATAGG ACATGATTTA GGGTTGGATT   2790
TGGAAAGGAG AAGAAAGTAG AAAAGATGAT GCCTACATTT TTCACTTAGG CAATTTGTAC   2850
CATTCAGTGA AATAGGGAAC ACAGGAGGAA GAGCAGGTTT TGGTGTATAC AAAGAGGAGG   2910
ATGGATGACG CATTCGTTT TGGATCTGAG ATGTCTGTGG AACGTCCTAG TGGAGATGTC    2970
CACAAACTCT TCTACATGTG GTTCTGAGTT CAGGACACAG ATTTGGGCTG GAGATAGAGA   3030
TATTGTAGGC TTATACATAG AAATGGCATT TGAATCTATA GAGATAAAAA GACACATCAG   3090
AGGAAATGTG TAAAGTGAGA GAGGAAAAGC CAAGTACTGT GCTGGGGGGA ATACCTACAT   3150
TTAAAGGATG CAGTAGAAAG AAGCTAATAA ACAACAGAGA GCAGACTAAC CAAAAGGGGA   3210
GAAGAAAAAC CAAGAGAATT CCATCGACTC CCAGGAGAGC ATTTCAAGAT TGAGGGGATA   3270
GGTGTTGTGT TGAATTTTGC AGCCTTGAGA ATCAAGGGCC AGAATACAGC TTTTAGATTT   3330
AGCAACAAGG AGTTTGGTGA TCTTCAGTGAA AGCAGCTTGA TGGTGAAATG GAGGCAGAGG  3390
CAGATTGCAA TGAGTGAAAC AGTGAATGGG AAGTGAAGAA ATGATACAGA TAATTCTTGC   3450
TAAAAGCTTG GCTGTTAAAA GGAGGAGAGA AACAAGACTA GCTGCAAAGT GAGATTGGGT   3510
TGATGGAGCA GTTTTAAATC TCAAAATAAA GAGCTTTGTG CTTTTTTGAT TATGAAAATA   3570
ATGTGTTAAT TGTAACTAAT TGAGGCAATG AAAAAAGATA ATAATATGAA AGATAAAAAT   3630
ATAAAAACCA CCCAGAAATA ATGATAGCTA CCATTTTGAT ACAATATTTC TACACTCCTT   3690
TCTATGTATA TATACAGACA CAGAAATGCT TATATTTTTA TTAAAAGGGA TTGTACTATA   3750
CCTAAGCTGC TTTTTCTAGT TAGTGATATA TATGGACATC TCTCCATGGC AACGAGTAAT   3810
TGCAGTTATA TTAAGTTCAT GATATTTCAC AATAAGGGCA TATCTTTGCC CTTTTTATTT   3870
AATCAATTCT TAATTGGTGA ATGTTTGTTT CCAGTTTGTT GTTGTTATTA ACAATGTTCC   3930
CATAAGCATT CCTGTACATC AATGTTCACA CATTTGCTG ATTTTTTCTT CAGGATAAAA    3990
CCCAGGAGGT AGAATTGCTG GGTTGATAGA AGAGAAAGGA TGATTGCCAA ATTAAAGCTT   4050
CAGTAGAGGG TACATGCCGA GCACAAATGG GATCAGCCCT AGATACCAGA AATGGCACTT   4110
TCTCATTTCC CCTTGGGACA AAAGGGAGAG AGGCAATAAC TGTGCTGCCA GAGTTAAATT   4170
TGTACGTGGA GTAGCAGGAA ATCATTTGCT GAAAATGAAA ACAGAGATGA TGTTGTAGAG   4230
GTCCTGAAGA GAGCAAAGAA AATTTGAAAT TGCGGCTATC AGCTATGGAA GAGAGTGCTG   4290
AACTGGAAAA CAAAAGAAGT ATTGACAATT GGTATGCTTG TAATGGCACC GATTTGAACG   4350
CTTGTGCCAT TGTTCACCAG CAGCACTCAG CAGCCAAGTT TGGAGTTTTG TAGCAGAAAG   4410
ATAAATAAGT TAGGGATTTA ATATTCTGGC CAAATGGTAG ACAAAATGAA CTCTGAGATC   4470
CAGCTGCACA GGGAAGGAAG GGAAGACGGG AAGAGGTTAG ATAGGAAATA CAAGAGTCAG   4530
GAGACTGGAA GATGTTGTGA TATTTAAGAA CACATAGAGT TGGAGTAAAA GTGTAAGAAA   4590
ACTAGAAGGG TAAGAGACCG GTCAGAAAGT AGGCTATTTG AAGTAACAC TTCAGAGGCA    4650
CAGTAGTTCT GAATGGTAAT AAGAAATTGA AGAAGCCTTT GAGAGTAGGT TAAAAACAA    4710
TAGGCAACTT TATTGTAGCT ACTTCTGGAA CAGAAGATTG TCATTAATAG TTTTAGAAAA   4770
CTAAAATATA TAGCATACTT ATTTGTCAAT TAACAAAGAA ACTATGTATT TTTAAATGAG   4830
ATTTAATGTT TATTGTAG  AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT    4880
                    Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu
                                                                -5                           1                              5
GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC                             4928
Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe
                10                           15                           20
ATT GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC                             4976
Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp
                25                           30                           35
TGT AGA G   GTATTTTTT TTAATTGGCA AACATAGAAA TGACTAGCTA CTTCTTCCCA                            5032
Cys Arg Asp
          40
TTCTGTTTTA CTGCTTACAT TGTTCCGTGC TAGTCCCAAT CCTCAGATGA AAAGTCACAG   5092
GAGTGACAAT AATTTCACTT ACAGGAAACT TTATAAGGCA TCCACGTTTT TTAGTTGGGG   5152
```

- 43 -

```
TAAAAAATTG GATACAATAA GACATTGCTA GGGGTCATGC CTCTCTGAGC CTGCCTTTGA   5212
ATCACCAATC CCTTTATTGT GATTGCATTA ACTGTTTAAA ACCTCTATAG TTGGATGCTT   5272
AATCCCTGCT TGTTACAGCT GAAAATGCTG ATAGTTTACC AGGTGTGGTG GCATCTATCT   5332
GTAATCCTAG CTACTTGGGA GGCTCAAGCA GGAGGATTGC TTGAGGCCAG GACTTTGAGG   5392
CTGTAGTACA CTGTGATCCT ACCTGTGAAT AGCCACTGCA CTCCAGCCTG GGTGATATAC   5452
AGACCTTGTC TCTAAAATTA AAAAAAAAAA AAAAAAAAAC CTTAGGAAAG GAAATTGATC   5512
AAGTCTACTG TGCCTTCCAA AACATGAATT CCAAATATCA AAGTTAGGCT GAGTTGAAGC   5572
AGTGAATGTG CATTCTTTAA AAATACTGAA TACTTACCTT AACATATATT TTAAATATTT   5632
TATTTAGCAT TTAAAAGTTA AAAACAATCT TTTAGAATTC ATATCTTTAA AATACTCAAA   5692
AAAGTTGCAG CGTGTGTGTT GTAATACACA TTAAACTGTG GGGTTGTTTG TTTGTTTGAG   5752
ATGCAGTTTC ACTCTGTCAC CCAGGCTGAA GTGCAGTGCA GTGCAGTGGT GTGATCTCGG   5812
CTCACTACAA CCTCCACCTC CCACGTTCAA GCGATTCTCA TGCCTCAGTC TCCCGAGTAG   5872
GTGGATTAC AGGCATGCAC CACTTACACC CGGCTAATTT TTGTATTTTT AGTAGAGCTG   5932
GGGTTTCACC ATGTTGGCCA GGCTGGTCTC AAACCCCTAA CCTCAAGTGA TCTGCCTGCC   5992
TCAGCCTCCC AAACAAACAA ACAACCCCAC AGTTTAATAT GTGTTACAAC ACACATGCTG   6052
CAACTTTTAT GAGTATTTTA ATGATATAGA TTATAAAAGG TTGTTTTTAA CTTTTAAATG   6112
CTGGGATTAC AGGCATGAGC CACTGTGCCA GGCCTGAACT GTGTTTTTAA AAATGTCTGA   6172
CCAGCTGTAC ATAGTCTCCT GCAGACTGGC CAAGTCTCAA AGTGGGAACA GGTGTATTAA   6232
GGACTATCCT TTGGTTAAAT TTCGCCAAAT GTTCCTGTGC AAGAATTCTT CTAACTAGAG   6292
TTCTCATTTA TTATATTTAT TTCAG AT AAT GCA CCC CGG ACC ATA TTT ATT    6343
                            Asp Asn Ala Pro Arg Thr Ile Phe Ile
                             40                    45
ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC   6391
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
     50                  55                  60
TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT   6439
Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile
 65              70                  75                  80
ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTTT CAATCATGTT AATATAATCA   6495
Ile Ser Phe Lys
ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA AGTAATGTAA TTAGAAAACT   6555
CAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAATA ACAAGAAGCA GAGAACCATT   6616
AAAGTGAATA CTTACTAAAA ATTATCAAAC TCTTTACCTA TTGTGATAAT GATGGTTTTT   6675
CTGAGCCTGT CACAGGGGAA GAGGAGATAC AACACTTGTT TTATGACCTG CATCTCCTGA   6735
ACAATCAGTC TTTATACAAA TAATAATGTA GAATACATAT GTGAGTTATA CATTTAAGAA   6795
TAACATGTGA CTTTCCAGAA TGAGTTCTGC TATGAAGAAT GAAGCTAATT ATCCTTCTAT   6856
ATTTCTACAC CTTTGTAAAT TATGATAATA TTTTAATCCC TAGTTGTTTT GTTGCTGATC   6916
CTTAGCCTAA GTCTTAGACA CAAGCTTCAG CTTCCAGTTG ATGTATGTTA TTTTTAATGT   6976
TAATCTAATT GAATAAAAGT TATGAGATCA GCTGTAAAAG TAATGCTATA ATTATCTTCA   7036
AGCCAGGTAT AAAGTATTTC TGGCCTCTAC TTTTTCTCTA TTATTCTCCA TTATTATTCT   7096
CTATTATTTT TCTCTATTTC CTCCATTATT GTTAGATAAA CCACAATTAA CTATAGCTAC   7156
AGACTGAGCC AGTAAGAGTA GGCAGGGATG CTTACAAATT GGCAATGCTT CAGAGGAGAA   7216
TTCCATGTCA TGAAGACTCT TTTTGAGTGG AGATTGCCA ATAAATATCC GCTTTCATGC   7276
CCACCCAGTC CCCACTGAAA GACAGTTAGG ATATGACCTT AGTGAAGGTA CCAAGGGGCA   7336
ACTTGGTAGG GAGAAAAAAG CCACTCTAAA ATATAATCCA AGTAAGAACA GTGCATATGC   7396
AACAGATACA GCCCCCAGAC AAATCCCTCA GCTATCTCCC TCCAACCAGA GTGCCACCCC   7456
TTCAGGTGAC AATTTGGAGT CCCCATTCTA GACCTGACAG GCAGCTTAGT TATCAAAATA   7516
GCATAAGAGG CCTGGGATGG AAGGGTAGGG TGGAAGGGT TAAGCATGCT GTTACTGAAC   7576
AACATAATTA GAAGGGAAGG AGATGGCCAA GGTCAAGCTA TGTGGGATAG AGGAAAACTC   7636
AGCTGCAGAG GCAGATTCAG AAACTGGGAT AAGTCGGAAC CTACAGGTGG ATTCTTGTTG   7696
AGGGGACTCC GTGAAGGATG TAAGAAGATG GAAATAATGC TTGGCACTTA GTAGGAACTG   7755
GGCAAATCCA TATTTGGGGG AGCCTGAGT TTATTCAATT TTGATGGCCC TTTTAAATAA   7816
AAAGAATGTG GCTGGGCGTG GTGGCTCACA CCTGTAATCC CAGCACTTTG GGAGGCCGAG   7876
GGGGCGGAT CACCTGAAGT CAGGAGTTCA AGATCAGCCT GACCAACATG GAGAAACCCC   7935
ATCTCTACTA AAAATACAAA ATTAGCTGGG CGTGGTGGCA TATGCCTGTA ATCCCAGCTA   7995
CTCGGGAGGC TGAGGCAGGA GAATCTTTTG AACCCGGGAG GCAGAGGTTG CGATGAGCCT   8055
AGATCGTGCC ATTGCACTCC AGTCTGGGCA ACAAGAGCAA AACTCCGTCT CAAAAAAAAA   8116
AAAAAAAAG TGAAATTAAC CAAAGGCATT AGCTTAATAA TTTAATACTG TTTTTAAGTA   8176
GGGCGGGGG TGGCTGGAAG AGATCTGTGT AAATGAGGGA ATCTGACATT TAAGCTTCAT   8236
CAGCATCATA GCAAATCTGC TTCTGGAAGG AACTCAATAA ATATTAGTTG GAGGGGGGA   8296
GAGAGTGAGG GGTGGACTAG GACCAGTTTT AGCCCTTGTC TTTAATCCCT TTTCCTGCCA   8356
CTAATAAGGA TCTTAGCAGT GGTTATAAAA GTGGCCTAGG TTCTAGATAA TAAGATACAA   8416
CAGGCCAGGC ACAGTGGCTC ATGCCTATAA TCCCAGCACT TTGGGAGGCC AAGGCGAGTG   8476
TCTCACTTGA GATCAGGGAGT TCAAGACCAG CCTGGCCAGC ATGGCGATAC TCTGTCTCTA   8536
CTAAAAAAAA TACAAAAATT AGCCAGGCAT GGTGGCATGC ACCTGTAATC CCAGCTACTC   8596
GTGAGCCTGA GGCAGAAGAA TCGCTTGAAA CCAGGAGGTG TAGGCTGCAG TGAGCTGAGA   8656
TCGCACCACT GCACTCCAGC CTGGGCGACA GAATGAGACT TTGTCTCAAA AAAGAAAAA   8716
```

- 44 -

```
GATACAACAG GCTACCCTTA TGTGCTCACC TTTCACTGTT GATTACTAGC TATAAAGTCC    8776
TATAAAGTTC TTTGGTCAAG AACCTTGACA ACACTAAGAG GGATTTGCTT TGAGAGGTTA    8836
CTGTCAGAGT CTGTTTCATA TATATACATA TACATGTATA TATGTATCTA TATCCAGGCT    8896
TGGCCAGGGT TCCCTCAGAC TTTCCAGTGC ACTTGGGAGA TGTTAGGTCA ATATCAACTT    8956
TCCCTGGATT CAGATTCAAC CCCTTCTGAT GTAAAAAAAA AAAAAAAAAA GAAAGAAATC    9016
CCTTTCCCCT TGGAGCACTC AAGTTTCACC AGGTGGGGCT TTCCAAGTTG GGGGTTCTCC    9076
AAGGTCATTG GGATTGCTTT CACATCCATT TGCTATGTAC CTTCCCTATG ATGGCTGGGA    9136
GTGGTCAACA TCAAAACTAG GAAAGCTACT GCCCAAGGAT GTCCTTACCT CTATTCTGAA    9195
ATGTGCAATA AGTGTGATTA AAGAGATTGC CTGTTCTACC TATCCACACT CTCGCTTTCA    9255
ACTGTAACTT TCTTTTTTTC TTTTTTTCTT TTTTCTTTT TTTTTGAAAC GGAGTCTCGC    9316
TCTGTCGCCC AGGCTAGAGT GCAGTGGCAC GATCTCAGCT CACTGCAAGC TCTGCCTCCC    9376
GGGTTCACGC CATTCTCCTG CCTCACCCTC CCAAGCAGCT GGGACTACAG GCGCCTGCCA    9436
CCATGCCCAG CTAATTTTTT GTATTTTTAG TAGAGACGGG GTTTCACCGT GTTAGCCAGG    9496
ATGGTCTCGA TCTCCTGAAC TTGTGATCCG CCCGCCTCAG CCTCCCAAAG TGCTGGGATT    9556
ACAGGCGTGA GCCATCGCAC CCGGCTCAAC TGTAACTTTC TATCTGGTT CATCTTCCCC    9616
TGTAATGTTA CTAGAGCTTT TGAAGTTTTG GCTATGGATT ATTTCTCATT TATACATTAG    9676
ATTTCAGATT AGTTCCAAAT TGATGCCCAC AGCTTAGGGT CTCTTCCTAA ATTGTATATT    9736
GTAGACAGCT GCAGAAGTGG GTGCCAATAG GGGAACTAGT TTATACTTTC ATCAACTTAG    9795
GACCCACACT TGTTGATAAA GAACAAAGGT CAAGAGTTAT GACTACTGAT TCCACAACTG    9856
ATTGAGAAGT TGGAGATAAC CCCGTGACCT CTGCCATCCA GAGTCTTTCA GGCATCTTTG    9916
AAGGATGAAG AAATGCTATT TTAATTTTGG AGGTTTCTCT ATCAGTGCTT AGGATCATGG    9976
GAATCTGTGC TGCCATGAGG CCAAAATTAA GTCCAAAACA TCTACTGGTT CCAGGATTAA    10036
CATGGAAGAA CCTTAGGTGG TGCCCACATG TTCTGATCCA TCCTGCAAAA TAGACATGCT    10096
GCACTAACAG GAAAAGTGCA GGCAGCACTA CCAGTTGGAT AACCTGCAAG ATTATAGTTT    10156
CAAGTAATCT AACCATTTCT CACAAGGCCC TATTCTGTGA CTGAAACATA CAAGAATCTG    10216
CATTTGGCCT TCTAAGGCAG GGCCCAGCCA AGGAGACCAT ATTCAGGACA GAAATTCAAG    10276
ACTACTATGG AACTGGAGTG CTTGGCAGGG AAGACAGAGT CAAGGACTGC CAACTGAGCC    10336
AATACAGCAG GCTTACACAG GAACCCCACTG CCTAGCCCTA CAACAATTAT TGGGTCTATT    10396
CACTGTAAGT TTTAATTTCA GGCTCCACTG AAAGAGTAAG CTAAGATTCC TGGCACTTTC    10456
TGTCTCTCTC ACAGTTGGCT CAGAAATGAG AACTGGTCAG GCCAGGCATG GTGGCTTACA    10516
CCTGGAATCC CAGCACTTTG GGAGGCCGAA GTGGGAGGGT CACTTGAGGC CAGGAGTTCA    10576
GGACCAGCTT AGGCAACAAA GTGAGATACC CCCTGACCCC TTCTCTACAA AAATAAATTT    10636
TAAAAATTAG CCAAATGTGG TGGTGTATAC TTACAGTCCC AGCTACTCAG GAGGCTGAGG    10696
CAGGGGGATT GCTTGAGCCC AGGAATTCAA GGCTGCAGTG AGCTATGATT TCACCACTGC    10756
ACTTCTGGCT GGGCAACAGA GCGAGACCCT GTCTCAAAGC AAAAAGAAAA AGAAACTAGA    10816
ACTAGCCTAA GTTTGTGGGA GGAGGTCATC ATCGTCTTTA GCCGTGAATG GTTATTATAG    10876
AGGACAGAAA TTGACATTAG CCCAAAAAGC TTGTGGTCTT TGCTGGAACT CTACTTAATC    10936
TTGAGCAAAT GTGGACACCA CTCAATGGGA GAGGAGAGAA GTAAGCTGTT TGATGTATAG    10996
GGGAAAACTA GAGGCCTGGA ACTGAATATG CATCCCATGA CAGGGAGAAT AGGAGATTCG    11056
GAGTTAAGAA GGAGAGGAGA TCAGTGACTGC TGTTCAGAGA TTTTTTTTAT GTAACTCTTG    11116
AGAAGCAAAA CTACTTTTGT TCTGTTTGGT AAATATACTTC AAAACAAACT TCATATATTC    11176
AAATTGTTCA TGTCCTGAAA TAATTAGGTA ATGTTTTTTT CTCTATAG GAA ATG AAT    11233
                                                        Glu Met Asn
                                                             85
CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG    11281
Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Glu
         90                   95                  100
AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA    11329
Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser
        105                  110                  115
TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA    11377
Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys
120                  125                  130                  135
CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC    11425
Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe
                140                  145                  150
ACT GTT CAA AAC GAA GAT TAGCTATTAA AATTTCATGC C                    11464
Thr Val Gln Asn Glu Asp
            155
```

(2) INFORMATION FOR SEQ ID NO:14:

    (1) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 28994 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: human
    (F) TISSUE TYPE: placenta

(ix) FEATURE:
    (A) NAME/KEY: 5'UTR
    (B) LOCATION: 1..15606
    (C) IDENTIFICATION METHODS: E
    (A) NAME/KEY: leader peptide
    (B) LOCATION: 15607..15685
    (C) IDENTIFICATION METHODS: S
    (A) NAME/KEY: intron
    (B) LOCATION: 15686..17056
    (C) IDENTIFICATION METHODS: E
    (A) NAME/KEY: leader peptide
    (B) LOCATION: 17057..17068
    (C) IDENTIFICATION METHODS: S
    (A) NAME/KEY: intron
    (B) LOCATION: 17069..20451
    (C) IDENTIFICATION METHODS: E
    (A) NAME/KEY: leader peptide
    (B) LOCATION: 20452..20468
    (C) IDENTIFICATION METHODS: S
    (A) NAME/KEY: mat peptide
    (B) LOCATION: 20469..20586
    (C) IDENTIFICATION METHODS: S
    (A) NAME/KEY: intron
    (B) LOCATION: 20587..21920
    (C) IDENTIFICATION METHODS: E
    (A) NAME/KEY: mat peptide
    (B) LOCATION: 21921..22054
    (C) IDENTIFICATION METHODS: S
    (A) NAME/KEY: intron
    (B) LOCATION: 22055..26827
    (C) IDENTIFICATION METHODS: E
    (A) NAME/KEY: mat peptide
    (B) LOCATION: 26828..27046
    (C) IDENTIFICATION METHODS: S
    (A) NAME/KEY: 3'UTR
    (B) LOCATION: 27047..28994
    (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
ACTTGCCTTA AAAGCTTTGC ATAGGTAGAC AACATTAGAT TAATTTCCTT GCTCACATCT      60
GTTCAAGAAA AATCATTTAA GTTATAAAAT ATAACAAACC TTCTGCATTA TAAGACTGAT     120
GTTTAGAAAT ATAAACATTT TATACATCAC CATTTAAATC TTTCTCCAAG GCTTCATCTT     180
TATAAAATAG TCCGGAAATT TCAGAGAAAG ATGAATCTGA TTTTCCAAGA GAGGACAGCT     240
GTGGACTATC TGGCACTGGA GACTAAATAA AGAAAGCAGG TACAGTCAAT AAGATCTTCA     300
GGACATATAC ATTTTGTTTA TTAAGAAAAA GCAAATAAAA CATTTTTCAG AAAAAGGCAA     360
ACATGCTAGA AAGCATATGA CTTAGTCATT TGAGTTTTTA TTATTAAGGA AATTTACAGG     420
CCCAAGAAAC ACCTTGCTCA ATATATTAAA TTTTATTTTG GTTTTCAACT AGACTTTGCT     480
TTTCATTTGT TTGTTTTTGT GACAAGTTCT CGCTCTGTCA CCTAGGCCAA AGTGTAGTGA     540
CACAATCTTA GCTCACTGTA GCCTCCTAGA TTCAAGTGAT CCTCCTGTCT CAGACTCCTG     600
AGTAGCTAGG ACTACAGGAA CATTCCACCA TGCCCAGCTA ATTTTGTTTT GTTTTGTTTT     660
GTTTTCAGAG ACAATGTATT GCAGCGTTGC CCAGGCTGAT CTGAAACTCT TAGCCTCAAA     720
CGATACTCCT GCCTCAGCCT CCCAAAGCAC TAGGATTACA GACATGAGCC AATGCGCCCA     780
GCCTTAAATT AGACTTTAAA TGTGGTTTTA AACTCCTGTT GAAAAAGCGT CTGGTATCTT     840
GAACCAGTAG ATGTTTTCAT AGCAATGAAG CTAAACTGTA ATTTAGACAG TAGCCAAATG     900
CTTGTGAAAT TTTGCTAAAT AATATAATCT TCAAGGGAGC AAATCATGTC CCAAATGCAA     960
AAGATCAACT GGTGGGGGCA GTAGTAAAAG ACAGGATACT GTGCTCTTTA AAAGGTCAGT    1020
AACTATAGTA CCTAGTTATC TTACTTATCA CAGCAAAATA ATTACATAAA ATCCTATGGA    1080
TCATAAAGGC ACAGACTCAC TTCTGTCTCT AGATCTCAAG CTACCAAAAA GAAATCTCCC    1140
AATAGTTTCT TGGAGGCCTA TACTTAGTGA AAAAGCAGCT GGAATCAACA TAGTTCCTCC    1200
TATGTTGTAG GACAATCCTA GCTCTGGGCA TACGAATACA TTAAATCCCA CTTATCTATA    1260
```

- 46 -

```
GAGCTTTCTT AAAGGGAAGA AATTTGAGTA GTATGTAAAA CAGAATAAAA GATTAAGGCT 1320
CCATAGGCAT ACAGCTTACC TCCAATTCTC TTGGCCTCTT GCAATTTCTA TTATCAGGCT 1380
TTACAAGGTG ATTTGCCATC ATATTCCGAA GGCACCAGCT ACAAAGCTTA GAACAATGCC 1440
AGATTTAGGT ACAAACTCCA TGCTACAAGC TCTCTGGAAT CCTTCCCTGT TTCCCACTCC 1500
TACTGCTGAT GTTAATTTAG ACTGTCATTA TCTGTCACTT TCCTAAACTC AATTCTCCC 1560
TCCTCTAAAT CATTCTATCA ACTGCTATTT GGGTAATCTT TCAAAACTTT GATTACTCCA 1620
TTCCTTTAAC TCAAAAACTT TCATTGTTCC AGAATAAGTT GAAATTCCAT GATATGGCCT 1680
TCAAGGTCCT GTATTATCTG GTGCAAGCCT ACTAGTCCCA TCATTTTCAA CTACTCCTCT 1740
CTATGTACTT AGCCAAATGA GTCTCTCTGG CAATTCTGCC TTGTTTCAGG ACTGGCTCAG 1800
TTAAGATTCT TTTATCTTCG GCCGGGCGCG CTGGCTCACG GCTGTAATCC CAGCACTTTG 1860
GGAAGCTGAG GCAGGAAGAT CACCTGAGGT CGGGAGTTCG AGACCAGCCT GGCCAGCATG 1920
GTGAAACCCT GTGTCTACTA AAAATCCAAA CATTAGCCAG GCGTGGTGGC AGGCGCCTGT 1980
AATCCCAGCT ACTTGGGAAG CTGAGGTGAG AGAATCGCTT GAACCCAGGA GAGGGAGGTT 2040
GCAGTGAGCC GAGATTGTGC CATTGCACTC CAGCCTGGGC AACAGAGCGA GACTCCACCT 2100
CAAAAAAAAA AAGGATTCTT CTATCTTCAC AAAATCTTAA TGTTTAAACA GGTCTTACAG 2160
TTCATCTAAT TCAATCTCAT TTTTTACAAG TGAGAAAACA GGGACAGTGA CGGTGGATCA 2220
AGTGACACCA GTAAGACTGA GCTAAATTAG AACCGAGATC TCACTCGAGT CTGAGGTTAT 2280
TCCCACTGTC CAACCTTACT TTAAAGTAGG TTCAAATTTT ACTTTTACTT TTCCATAAAT 2340
TCGGAAGGGA TTTTCCCTAG GAGTCCAAAT GTTGAAACCT GGAAGGGTAT AGTCTCTGTG 2400
TCTTTGAGAT GAGGGGAGCC CTGTCCATAT TCAAGTTATC AATTGACTTT GTTGTTTTTG 2460
AGAAACGATG CTGATTTGGG TAACTTTAAC ACATCTGTTT GATTAGTCCT ATAAAATATG 2520
CATATATAGA AGACAGAAAG AGCAACAACA AATTTGAAAG ATGCTTGTTA AGTAAATTCT 2580
GTATCGTACG TGTCCATTCC TGCCAGTACC TTTATAGTAT GTAAGTTTAC GTGCTGTAAT 2640
AGTATTAATA GTATCTAGAA AATACTACAC ATGCACAGCA GTGCTAACTT TGCCTTGGGA 2700
GTTGGAAAAT ACTTCAGAGA AGCCAACAGG CAGATTTTTC TCTCTTCCCT TCCCCTTCTA 2760
ATTTTCCCTT TCCCCTTCAC CCCCTTCTCT TCTCTCCCCA AGTAACACTG TGCACCTATG 2820
TCAAACGAAA ACTTATAATC AAGTAACTGT TTCTGCAAAA ATAAGTTCGT TTTCCTGTCA 2880
TGGCTCAAGG CCTCAGCAGA TCCAGGCCTG GTGGACGGGC TGGTCTTCGT CGTGTGCCAA 2940
ACACTGACCA CTGCCCTGGC TCTGCCATCT TAGGCTTAGT GACCTGGCTG TTACTAAGCA 3000
CTGTCCCCTC TGCCCCACGA AGCTGTCTCC TTCTTAGTCTT CTCCCTCTTC TCAACGCGAT 3060
CCTAGCCCCT CAGGCCATTT CACCTCCATT TTCCCTCACT TCCCGCCCGC CCTCCGCACT 3120
TCCTCCCTAC TGTTGTTTCC GCCCCACTAG AGCCCCTCAG AGAAAGTTTC CATCCTCGCA 3180
CCCTTCCTTG TGTCACAGCC CGTCACATTC TCACAGGCGC CCATCCCTCC AGCCCCACCC 3240
CAAGGCCAAT GTACTTCGCG GTATGGGGAC CTTCCTCGTC AGCGAACGCG AGGGAGTGAA 3300
GACCCTGGGC GCGGGGTGCT CGGACTTCGG GGGTGGAGGT GGGAAGCGCG CCGCACTCCC 3360
AGCAGCCCCT GCACGAGTCA CGTGACAGCT CTCCCACCAC CAGCCCCCCC AACTTCCCCA 3420
CCGTAGCCTC CCAGAGCCAG GCCCCACGGA AAGGCAGCTT TTTCCCGGTT TTCTCCCGCT 3480
CTTTCCCCTC CACTTGGAAT ACTCGTGAAA CAAAAATCTC TCCCTGCCAC CCTGTGTGTG 3540
TTTGAATCAG GAAAAAATCT GAAACTGGTC AAGAAAGAAC AAGGAAGACT TGCCAAAGCA 3600
AGGCCGGTGT GTGTCCCAGC AGCTTAGAAT CTCAGCAAAG GAACACAAAA TAGCACATCC 3660
AGGGCCTCTT TTCGGACAAA ATTTACTTGG TTTGTTTGCA GGAAGGGTTT AAAACTGCGT 3720
TTGCAGATGC TCTGTTTGCA GGAAGGCTTT AATCACGTGT TCCCCTGGGC CACAAGCAAG 3780
GCTTTTAGAT CCAGAGCCTC AGTTACTGCC CCCTCTTCCT CTTTGGTGCA ACCAAACGTT 3840
CAGAATCACG CCTTCTTAGA AAATTCTTAC CCCGGGTGTG TCAATAAGTT AAGTCTAATT 3900
GGCAACAGCT ATCAAAAAGT GTTGCATAAC ACACATGGCT CACATAATTG TAGCTTTGCC 3960
TCATCGGGTG TTTTAATGCG GAGGCTTTGA CCTGCAATTT CAAAGATATA CATTCCAAGC 4020
TTACGCCCAG TTAGTGGATG TGGAAGAAAA AAAAAAGCAA ATTACCTCAT AACACAAAGG 4080
TCAATAACAC ACATCCATAA GTCCAGGTA CAAAATCTTA CATCTTAGAG AACTATATTT 4140
AACATTTACA TACATTACTA AGGTTTTTTT TTTCCTTTTG CTTGATTAAA TGTTAGTTAT 4200
CATTAAGTCT TGGAATTATT CTGTGTGTGT ATATTTATTT GCTGTTTGTG AAGAAGCCGG 4260
TTGTTTTAAA TAAGTTCCTA GAAAATAAGC GCTCAATGTG TTTAATCTGA GTTGCTAATA 4320
TTGTGAAATA TAGGCCACAT AATACTAGCC TACATAACTA TGGCGAAATA AGGAGTCTCA 4380
AACACTGCTA CAGAACAATA GCAATCTGTG TTGAATTTTT ACCCTCTGTG GTAAAATGAA 4440
GGGAAAAGGA ATGAAGTTTT AGTTTGCCTTT AATTTTTATC TTTATTGTTT CAGACTCTTC 4500
AGCAGTATAA AGTTTTCATC AAGTCAAATA TATTCACTTT AAAGTGACTG TGCTTTATTC 4560
TGATACCATG TCCTTCCTAA TTTGGGGGGC CAGGTGAGAT AAGTTTTATG AAATAAAAAG 4620
ATTAAAAATT CTTACATTTT TAGTGTCCTT CCTTGGTAAA ATGTAGAGTT GTCCACTGTG 4680
TTTATCTCCT CCTCCTTATT ATCATGGTTG CTCTTATTAT TTTAATGGT TCATTAAACC 4740
CAAGGTCTG GGAAATACTC ATGGCATTCA TCTCACAGCC TTCACACTGT ATGATATTTA 4800
AACAGGTGGT TGTCCATCTG ATTCTTAAAA TATTTCCAAG AAAAATGATT CCACCTAATG 4860
CATAAATGCT TTCATCAGAT TAAGAGAACA CCATGGACAT TTTATTTTAT TTTATTTTTT 4920
AAATATTAAC TTCCATTGCA TAAGCTAAAT GGTAGGAAT AAGTGAGATG ATATTGTTAT 4980
CTAGAGCTTT AAAATATTCA AAGGGCTGTC ATCATTATCT CATTTAATCT TTGAAAACAA 5040
CTCTATGAAG TACAAAGGAC ACTGAGACAT TTGTTGCTCT ATATCAAAGA AAAAGTGTT 5100
TGTCCCAAAA CTTCAAAAGC TGTAAATTAC ACATTCACGA TCTTTACAGC TGGAGAAAAT 5160
TCACTGGCAA TGGAATATTT AAAATTAGAG CTTGCTTAGT GTGCTGCTTC TGATCACTAC 5220
TTGATCCCAC TTCGTGCTTT CATGTTAATT GGCCCAATTG GACTCTACAG TTGGAAGGTG 5280
```

```
AAAACTTACT ATTTCAACTT GAGTCACGTA TGTATTCTTA TCATATACTT CTTAAAGGTA      5340
CTATTTTTTT TCTTCTGATA GTCACCACAC CAAGCACTTC CAGCCACCCT GCCACAGACT      5400
TCCTTTGTAA TCACTGTTGA AGGACATGAT GTTTTTATGA CTTCCCGAAA TGAAAACCCT      5460
ATCTTGTTTT TAAAACAAAC AAACCAACAA AAAGTAGTGT TTATGTAAGC ATTTTGTTCC      5520
CTGACTCTAG GAACCCCTCT GTTTTTATAT CAACTCTGTA CTGGCAAAAC ACAAAAACAA      5580
AATGCCACCT TGCTAATTCC CTTCCTAGCA AAGTAATACA GTTAGCACA TGTTCAAGAA       5640
AAAAATGGCT AAGAAATTTT GTTTCCACTA ATTATTTTCA AGACTGTGAT ATTTACACTC      5700
TGCTCTTCAA ACGTTACATT TTATAAGACT ATTTTTTAAC ATGTTGAACA TAAGCCCTAA      5760
ATATATGTAT CCTTAAATTG TATTTCAAAT ATTTTAGGTC AGTCTTTGCT ATCATTCCAG      5820
GAATAGAAAG TTTTAACACT GGAAACTGCA AGTAAATATT TGCCCTCTTA CCTGAATTTT      5880
GGTAGCCCTC TCCCCAAGCT TACTTTCTGT TGCAGAAAGT GTAAAAATTA TTACATAAAA      5940
TTCTAATGAT GGTATCCGTG TGGCTTGCAT CTGATACAGC AGATAAAGAA GTTTTATGAA      6000
AATGGACTCC TGTTCCACTG AAAAGTAAAT CTTAATGGCC TGTATCAACT ATCCTTTGAC      6060
ACCATATTGA GCTTGGGAGG AAGGGGAAGT CCTGAATGAG GTTATAAAGT AAAAGAAAAT      6120
ATTTGCAAAA TGTTCCTTTT TTTAAAAATGT TACATTTTAG AAATATTTTA AGTGTTGTAA      6180
CATTGTAGGA ATTACCCCAA TAGGACTGAT TATTCCGCAT TGTAAAATAA GAAAAAGTTT      6240
TGTGCTGAAG TGTGACCAGG AAGTCTGAAA ATGAAGAGAG ACAGATGACA AAAGAAGATG      6300
CTTCTAATGG ACTAAGGAGG TGCTTTCTTA AAGTCAGAAA GAGATACTCA GAAAGAGGTA      6360
CAGGTTTTGG AAGGCACAGA GCCCCAACTT TTACGGAAGA AAAGATTTCA TGAAAATAGT      6420
GATATTACAT TAAAAGAAGT ACTCGTATCC TCTGCCACTT TATTTCGACT TCCATTGCCC      6480
TAGGAAAGAG CCTGTTTGAA GGCGGGCCCA AGGAGTGCCG ACAGCAGTCT CCTCCCTCCA      6540
CCTTCTTCCT CATTCTCTCC CCAGCTTGCT GAGCCCTTTG CTCCCCTGGC GACTGCCTGG      6600
ACAGTCAGCA AGGAATTGTC TCCCAGTGCA TTTTGCCCTC CTGGCTGCCA ACTCTGGCTG      6660
CTAAAGCGGC TGCCACCTGC TGCAGTCTAC ACAGCTTCGG GAAGAGGAAA GGAACCTCAG      6720
ACCTTCCAGA TCGCTTCCTC TCGCAACAAA CTATTTGTCG CAGGTAAGAA ATATCATTCC      6780
TCTTTATTTG GAAAGTCAGC CATGGCAATT AGAGGTAAAT AAGCTACAAA GCAATTGAGA      6840
GGAATATAAA CCATCAGCA TCACTACGAT GAGCAGTCAG TATCAACATA AGAAATATAA      6900
GCAAAGTCAG AGTAGAATTT TTTTCTTTTA TCAGATATGG GAGAGTATCA CTTTAGAGGA      6960
GAGGTTCTCA AACTTTTTGC TCTCATGTTC CCTTTACACT AAGCACATCA CATGTTAGCA      7020
TAAGTAACAT TTTTAATTAA AAATAACTAT GTACTTTTTT AACAACAAAA AAAAGCATAA      7080
AGAGTGACAC TTTTTTATTT TTACAAGTGT TTTAACTGGT TTAATAGAAG CCATATAGAT      7140
CTGCTGGATT CTCATCTGCT TTGCATTCAG ACTACTGCAA TATTGCACAG AATGCAGCCT      7200
CTGGTAAACT CTGTTGTACA CTCATGAGAG AATGGGTGAA AAAGACAAAT TACGTCTTAG      7250
AATTATTAGA AATAGCTTTC ACTTTAGGAA CTCCCTGAGA ATTGCTGCTT TAGAGTGGTA      7320
AGATAAATAA GCTTCTCTTT AAACGGAATC TCAAGACAGA ATCAGTTACA TTAAAAGCAA      7380
ACAAAAAATT TGCCCATGGT TAGTCATCTT GTGAAATCTG CCACACCTTT GGACTGGGCT      7440
ACAATTGGAT AATATAGCAT TCCCCGAGAT AATTTTCTCT CACAATTAAG GAAAGGGCTG      7500
AATAAATATC TCTGTTTGAA GTTGAATAAC AAAAATTAGG ACCCCTAAA TTTTAGGGCT      7560
CCTGAAATTC GTCTTTTTCC CTATATTCAG CTACTTTACG TTCTATTAAA TCTTCTTTCA      7620
GGCCAGGTGC ACTAGCTCAT GCCTAGAATC TCAGGCAGGC CTGAGCCCAG GAATTTGAGA      7680
CCAGCCAGGG CAACACAGTC TCTACAAAAA AATAAAAAAT TACCTGGGTG TGTTGGTGCA      7740
TGCCTGTAGA ACTACTCAGG ATGCTGAGGA CTGCTTGAGC CCAGGATAGC CAAATCTGTG      7800
GTGAGTTCAG CCACTAAACA GAGCGAGACT TTCTCAAAAA AATAAACAAA AAAACAAACA      7860
AACTTCCTTC AAAATAACTT TTTATCTGCA ATGTTTTCCT ATTGCCTGTG AGATTAAATT      7920
TACTCTTTTA CCTGATTTCC AAAGCCCTCC ATAATCTAAT CCGACTTTAC CTTGTGTTCA      7980
CTGCAAAATA GTAGGACTGT TCCACTACAA TCCAAAAATC ACAGGTTGGG TGCAGTGGCT      8040
CACTCCTGTA ATCCCAACAC TTTGGAAGGC CAAGGCAGGT GGATTGCTTC AGCTCAGGAG      8100
TTCAAGACCA GCCTGGGCAA CATGGCAAAA ACCCTGTCTC TCTAAAACAT ACAAAAATTA      8160
GCCAGATGTG GTAGTATGTG CCTGTAGTCC CAACTACTCA AAAGGCTAAG GTAAGAGGAT      8220
CACTTGAGCC CAGGAGGTCA AGGCTACATG TT TACTGTGTCA CTGCACTCCA      8280
GCCTGGGTGA TAGAGCAAGA CCATGTCTCA AAAAAAAAAA AAAGAAAAGA AAAGAAAAAA      8340
ACATCGCTCT ATTCAGTTCA CCCCCACCAC AACATTGTTT TGATTATCAC ATAAATGCTG      8400
GTCCATTGCC TTCTCTATCT ATTCAAATCT TTAAGCATTC TTTGAGATTC AACTCAATTC      8460
TCCTTTTCAA ACTAGGCCAT TTAAACTACA TCAGTTCCAT TTTGATTTTC TTGCTTTGAG      8520
TCTACAGCT CAAAAACAAA AACTTAAAAA CTTATTTTTT AAGTTTTCTG CTACTCTCAC      8580
TTCTTCAACA CTCACATACA CCCATTCATA ATAAGATGGC AGAATGTTCA AGGATAAAAT      8640
GATTATATGA ACTGAAAAGT TAGGTTTTGA TCTTGTTGCT GTCAAGATGA CTACCTACCT      8700
GATCTCAGGT AATTAATTAT GTAGCATGCT CCCTCATTTC ATCCCATACC TATTCAACAG      8760
GATTGGAATT CCACAGCAAG GATAAACATA ATCATACTTG CTTTTCAAGT TCAAGGCATT      8820
TTAACTTTTA ATCTAGTAGT ATGTTTGTTG TTGTTGTTGT TGTTTGAGAT GGAGCCTGC      8880
TGTGTCACCC AGGCTGGAGT GCAGTGGCAC GAACTCGGCT CACTGCCAAC TCTGCCTCAT      8940
GGGTTCAATC AGTTATTCTG CCTCAGTGTC CCAAGTAGTC TGGGACTACA GGCACATGCC      9000
ACCATGCCTG GCTAATTTTT GTATTTTTAG TAGAAACAGG GCTTCACCAT GTTGGCCAGG      9060
CTGGTCTCGA ACTCCTGACC TCAAGTGATC CAGCCGCCTC GGCTCCCAA AGTGCTGGGA      9120
TTACAGGCAT AAGCCACCGT GCCCAGCCTA ATAGTATGTT TTTAAACTCT TAGTGGCTTA      9180
ACAATGCTGG TTGTATAATA AATATGCCAT AAATATTTAC TGTCTTAGAA TTATGAAGAA      9240
GTGGTTACTA GGCCGTTTGC CACATATCAA TGGTTCTCTC CTTACAGCTT TAATTAGAGT      9300
```

```
CTAGAATTGC AGGTTGGTAG AGCTGGAACA GACCTTAAAG ATTGACTAGC CAACTTCCTT    9360
GTCCAAATGA GGGAACTGAG ACCCTTAAAA TTAAGTGACT TGCCCCAGAC AAAACTGGAA    9420
CTCATGTGTC CTAATTTCCA TCATGTAAATT CTACCATTCA CTACCCTCTG GCTAGTTGTC    9480
AAAGTATTGC ATAACTAAAT TTTTATGTCT GTTTTAAAGA ACAAATTGTC ACTGCTTACT    9540
CCTGGGAGGG TCTTTCTGAG GTGGTTTATA ACTCTTAAAA AAAAAAAAGT CAGTAGTCTG    9600
AGAATTTTAG ACGAAATAGT CAAAGTCATTT TTATCCAATG GATCTATAAT TTTCATAGAT    9660
TAGAGTTAAA TCAAAGAAAC AGGGATGAGA AAGGAAGAGG AAAATTGAGG AGAGGAGGAA    9720
TGGGATGAGG AACACACTAC TTGTAATCAG TCATAGATGT ACTGAGAACT AACAAGAAGA    9780
ATTGTAAGAA AATAAGAATG AAGAATTCAA AATCAACACA TGAAATAAAA AGAAACTACT    9840
AGGGAAAAAT GGAGAAGACA TTAGAAAAAT TATTCTATTT TTAAAATTCT GTTTTCAGGC    9900
TTCCCTCCTG TTCTTCCTCC TTCTCATTGG TTTTCAGGTG GAGGGAAAGT TTAAGATGGA    9960
AAAAATATAT ATATTCTACA CATCCCTTTC TACGCTGTTG TCATGGCAAC AAGGTTTATC    10020
ATAGCAAACT TTTATTCATA CAACATTTAT TGAGTTCTTA CTGTGTGGTA AGCTCTTTCC    10080
AGGTGTTGAA AATTCAGGGG AAAAAAGACA ACTCATTGTC TTAAAACTCA GATGAAAGCT    10140
GAACAGACCT ATTTTTAATC AAAGTAATCT CAATTTAGGG TAGTAAGAGC TATTTAAGAA    10200
GCATGAACAG GTGTGAAGGA GGTAGGACTC TGAGGAGAGA ATAGTTAGCT AGGAATGAAA    10260
GAGCAGAGAA GTTTTCCTAG AGGAACTATT AAAGCTGGGA GTTACGGGAT GAAAGATGAG    10320
GTAGGGTTTG CAGGCAAAAA AAAAAAAAAG GCAGGGGAAG GGGAAGTTCT GGCCTGGCAG    10380
AGAGAATAAC TGTGGCAACA ATGGAGGAGA GTCTGGAAGC AAGAAAACCA AGTAGAAGAG    10440
TATTAAAATA GAAGATGCCA GGGGTAATGA GGGCTTGATT TAAAACAGTG CTGTTGGAGA    10500
TGGAGAGGAG ATACCAAATT CTGGAGACAT TTCTGAGTTA GAACCTACAG TATTTATCAG    10560
ACAAGGGAAA GATTAGACAA AGGAGTTAAG AATGACTCCC AGGTTTCAGT TTGGGGCAGG    10620
TAACTAGGAC ATGTTTTGAA AAGTAATGTA TTGGATCTCT TACCATTGGA ACTATGTATG    10680
TGGAGCCAAA TTAAAATTTG TACATGTATA TAACTCTCCC CCCACCACCA GTAACTACTT    10740
CCCTAACTCT CTACTTTGTA GCCAGACTTC CTAAAAGAAT AGTTTGTAGT CACTGTCTTT    10800
ACTTTTCCCC TCCCATTCTG TCCTAGATAT TTGTCCACCT ACCATCTGCT GCCTCCACTT    10860
TACCCAAACT GTTCTACGGT TGCCCAAAAC TTCCTAATTG CCAAATTCAA TGAACAAGTT    10920
TAAGCTTATA TGTAAATTAG GAGCTCACAA GTTTGATTTC GAGCAGCCCC TCCTGAAACC    10980
CTTTCTCTTT CGACTTCTGT GACACATCTC AGATTTACAA AACTGAACTA ATTATTTTAC    11040
ACTTGAGCTG TATTTTCGTT CTTCTTTCTT GATGAATGAG GTAACCACTC AACAAATTGC    11100
CCAAGCCAAA AACTACGAAG TCATCCTCAG TTCCTCCTTC TTCTGTTTGA CCCACAACAG    11160
ATCAGCTGAG AAATCCCGCT GTTTAGTATC TCTTGAATTC ATTACCTTAA TTTATAGCCT    11220
CATCAACTCT TAATTGTTAA AATTACTTCA GTAGTTGTTG TCTGACCTCT GTCCAATCTT    11280
GTTCAATCAG GTCCATTCTT TTGTTCTTGG TGGTGGTGGT GGTGTTGACA GAGTTTCGCT    11340
TTTGCTGCCC AGGCTGAAGT GCAGTGGAGC ACTTCACTGC AACCACAGCC TCCTGGGTTT    11400
AAGCAGTTCA CCCTCCCGAG TAGCTGGGAC TACAGGTATG TGCCACCACA CCCAGCTAAT    11450
TTTGTGTTTT CAGTAGAGAC AGGGTTTCAC CATGTTGGTC AGGCTGGTCT CAAACTCCTG    11520
ACCTCAAGCA ATCCACCCAC CTCAGCCTCC CAAAGTGCTG GGATTACAGG CATGAGCCAC    11580
TGCACACGGA CCAGATCCAT TGTTTATGTT GCTTCTAGAG TGAGTTTTTA AAACACAAAT    11640
TTGACCATAT CTTTCTCCAA TTTAAGTCAA TATTTTTTTT TTCAGGAAAA AACAGTTCAA    11700
ACTCTTAGT CTGCTTACAC AAGGCCTTTG TAGTCTGACT CTTCTTTCCA AGCTTTCATC    11760
AAAGTATACT GCAAGTTACA TTTTATGTGA ATTGAATTAG GCAACGGTAT AAAAATTATA    11820
GTTTATATGG GCAAAATGGA AATAATGTTA ACTCTTCCAA ATAGTTATC TAGAATGACA    11880
TAATTTCAAA GCTGTCAGGT CAAATGAGTT ATAAACTGTT AACACTATTG CCACATGCAA    11940
GTGTCTCTTA TACTTGGTAG AATTATCTGC TTCCATGTCA TTATTATGTA AATTAGACTT    12000
TAAATAACTC AGAAGTTCTT CAGACATACA GGTTATTATT GTGCTTTTTA AACATAATTT    12060
TAAATAATTT TATATATGAT AATGTTATCC AAGTGCTAAG GGATGTATTG TTACTGCTGT    12120
GCAAAAAAAA AAAAAAAAAA AACTCCAAAT AAATATGTTG AAACCAAGTT TATATGCAAG    12180
AAAACAATAT TAAAAAGGCC AAAGTACCAC CATAATAGGC TGTGTGGAGA CGGCAGGCTA    12240
CAAAACACTA GTAATAATGC TGAGAAAGTT GAAAAGCAAC AGAAAGCAAC AATATGCTTT    12300
GGTTGTTGTA GGTTTATTGTA CTCCAAGAAT ATCTCCTCTC AAACTTTTAC GTTTTTTCCA    12360
AAGAAAAGTT AACTTTGGCT GGGCGCAGTG GCTCTTGCCT GTAGTCCCAG CCTTTGGGAG    12420
GCCAAGGCGG GCAGATCACC TGAGGTCAGG AGTTTGAGAC CAGCCTGACC AAAAATGGAG    12480
AAACCCCGCC CCCTCACTAC TAAAAGAATA CAAAATTAGG CGGGCACAG TGGCTTACGC    12540
CTGTGATCCC AGCACTTTGG GAGGCCGAGG CAGGAAGATC ACCTGAGGTC AGGAGTTGGA    12600
GACCAGCCAT GGAGAAACCC GTCTCTACTA AAAATACAAA ATTAGCCGGG CGTGGTGGTG    12660
CATGACTGTA ATCCCAGCTA CTCAGGAGGC TAAGGCAGAG AATCACTTGA ACCCAGGCAG    12720
TGGAGGTTGC AGTGAGCCGA GATCGTGCCA TTGCACTCCA GCCTGGGCAA CAAGAGCGAA    12780
ACTCTGTATC CAAAAAACAA AAGAAAAGAA AAGGTAACCT TGAACTATGT GAGATCTTTA    12840
GAAATGCATT CTTTCTGTAA AATGTGACTA CATTTGCCTT ATTTATGGTA AAAATGTTGA    12900
GGCCTCAAAC AACCCATATT TTCTGGGTCT CCCCGCTGCC TAGCCTTGT TCACATTGCT    12960
TCTTCTTGGT GGAAGCTCTT CCTCTGGCCT TGAAAATGCC TGCTTCTCTT TCAAGGTAGC    13020
ACAGTCATCA CTTTCTGTGG TAACCTTCTC CAGCACCATC AAACAGAAAG AATGAATCTC    13080
TTGTAAATTC AGCTCTTACG TCATTCATTA CATTATTTTG TAACTCTTTA TAGATTCTTC    13140
TCTCCCACTA GACTCTGAGT CACTGGAGAG TAGGAGCCAA CTCTCATTCA TGTGTGGTTT    13200
GGTCAGCTAC TGGCCACATT CCTGATGCAT AGTTAATGCT CAAATCTTAA CTGGTGAATC    13260
AGCTCAAATA TTGTCCTTCT CTAAATCCAT TCACTCATTG ACTAACTATG TACTCAAAAT    13320
```

```
AGTAAACACC AGTAATTTAA TCCAATTCCT GCCCATACTG CTTGGTACAT TTCAGGTGAA   13380
TTAGTTTGAT AAATATGTGT GTATTACATA ATATTAAAGT ATGTACAGAA GATCATGCTA   13440
ATCATAATTC ACAACTGATA ACTAATCAAA CATAAATGCT CTCAGGTTAA CAAATGTCTG   13500
CCTTCTCAGT TAATGCAGTC ATTAACAAAC ACCTTCTGAT GCTGATAATA GGGCCTTGTT   13560
CAGCAATGAA GCCATAAAGG TGAATAAAGA ACATGCCCTC GTGGAGCTCA CAGCCTAGTC   13620
ATTATTGTTC TGATTTTTAA TATTAATGTT GGTTTGGGTT TTGGTGAAAA ATGTTTAGAC   13680
TTATCTTAGT GATCTTTTCA TCCTTTGCTA TATTATTTTT CTCTAAGAGT CTTCCTTATC   13740
CCCTCCTTTA AAAAACTAGG TGATAATTCT AAATTGTAAA TTTAAATATT ATAAATAGCT   13800
TATAAAATTT AATATTTATA ATATTTAAAT GTTTGATAAA TATTTAAATT TTATAATATT   13860
TAAATGTTTA TTTAAATTCA TTTGTACATC AGTTTTTATT TTATTTAAAT GTGTTGGCCA   13920
GGCATGGTGG CTGACACCTA TAATCCCAGA ACTTTGAGAG GCCAAGTCAG GCAAACCATT   13980
TGAGCTCAGG AGTTTGAGAC CACCCTGGGC AACGTGGTGA AACCCTGTCT CTACCAAACA   14040
TATGAAAACT TATCTGGGTG TGGTGGCACG CATCTGTGGT CCCAGATGGG AGTCCCAGGC   14100
TAAGATGGGA GAATCGCTTG AACCCAGGTG AGAGGGGTGG GGTGGATGTT GCAGTGAGCT   14160
GAGATCGTGC CACTGCACTC CAACCTGGGT GACAGAGTGA GACTCCATCT CAAAAAAAAA   14220
AAATGTTATC TAAATAAGAT AAATTTAATA ACTGTTCGCA CTTAGATGAG CATAAGGAAC   14280
TAAACCTAGA TAAAACTATC AAATAAGGCC TGGGTACAGT GACTCATGCC TGTAATCTCA   14340
AGCACTTTGG GAGGCCAAAA TTATACAAAG TTAGTTGTAT AACACCAACT AACAACTATT   14400
TTGGGGTTAG CTTAATTCAG ATTAATTTTT TTTAAACTGA GTTTTAAATT CCTGCTTACT   14460
CTACCATACA TGCTAGGCCT CATATTATGC TAGAAAAATT TTGAGCACAG ATTTATGAAT   14520
ACTCTCCTGC ATACCATTTA ATTTTTAAAC AAATTTTAAT GCAGTATATA TGTGCCTTTT   14580
TACCAACACA TTAAATAATA AGATCTACTG TGAGGACTAA ATTTCTGTAA TTTCAAAGTA   14640
GTAATGAGTT TAAACCATGT CTCAAGATCT CTGCAATAAC TGTAGCACAA CAGAAAATAG   14700
GTATTTCTAT TAATGACAGA GTCACAAGTA CTACTAATAA TACTGTGGTT TGTTTCCTGC   14760
AACTAATCAT GGAGGAATG CTAAATTTCA GAGGTTGGTG AAAATACATG TGTATTTTTT   14820
TCCCCATCCA AGTTCACAGA TTTCTCACAC TGAGAACTCC TATTCCATAA CAAAATTCTG   14880
GAAGCCTGCA CACGTATTG GAAGAAGGGC AGAAAGGAAA AGCAAATGGA AGGATTTAAA   14940
TTTTTTTCAA ATCCTGTATC CCTTGATTTT ACAGCAAGAT TGTATTTATG TATTACTTGT   15000
GTTAAAAATA TAGTATAATA GAGACTCCA ATCAAAAATC ACGCAGCTC AGGGAGAAAG   15060
AGGGCCACCA AATGCCAGAG CCCTTCAGCC TTCTCCCACC CTGCCTGTAC CCTCAGATGG   15120
AAGCACTTTT TTATCATTGT TTCACCTTTA GCATTTTGAC AATGAAGTCA CAAACCTTCA   15180
GCCTCTCACC CATAGGAACC CACTGGTTGT AAGAGAAGGA TGAAGCCAGT CCTTCCTAAA   15240
GGGCACGATT AGATGTGTTT ATGGCATCCT CAGGTGAAAC TATATTTATA TTGACAATAT   15300
ATTTATATTT CTCAAGGAAT ACTAGAATAA TGATTCAGTT CAGTACTAGG CCATTTATCT   15360
ACCCTTTATA ATATTGTTTA ATGAGAAAAT GCTTCTATC TTCCAAATAT CTGATGATTT   15420
GTAAGAGAAC ACTTAAACAT GGGTATTCAT AAGCTGAAAC TTCTGGCATT TATTGAATGT   15480
CAAGATTGTT CATCAGTATA CTAGGTGATT AACTGACCAC TGAACTTGAA GGTAGTATAA   15540
AGTAGTAGTA AAAGGTACAA TCATTGTCTC TTAACAGATG GGTCTTGCT TTCATTAGGA   15600

ATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA   15651
       Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala
        -35               -30                 -25
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G   GTAAGGC TAATGCCATA   15702
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
 -20               -15                 -10
GAACAAATAC CAGGTTCAGA TAAATCTATT CAATTAGAAA AGATGTTGTG AGGTGAACTA   15762
TTAAGTGACT CTTTGTGTCA CCAAATTTCA CTGTAATATT AATGGCTCTT AAAAAAATAG   15822
TGGACCTCTA GAAATTAATC ACAACATGTC CAAGGTCTCA GCACCTTGTC ACACCACGTG   15882
TCCTGGTACT TTAATCAGCA GTAGCTCACT CTCCAGTTGG CAGTAAGTGC ACATCATGAA   15942
AATCCCAGTT TTCATGGGAA AATCCCAGTT TTCATTGGAT TTCCATGGGA AAAATCCCAG   16002
TACAAAACTG GGTGCATTCA GGAAATATAA TTTCCCAAAG CAAATTGGCA AATTATGTAA   16062
GAGATTCTCT AAATTTAGAG TTCGGTGAAT TATACTCATTT TATGTAAATA TGTTTGACAA   16122
GTAAAAATTG ATTCTTTTTT TTTTTTCTG TTGCCTAGGC TGGAGTGCAG TGGCACAATC   16182
TCTGCTCACT GCAACCTCCA CCTCCTGGGT TCAAGCAATT CTCCTGCCTC AGCCTTCTGA   16242
GTAGCTGGGA CTACAGGTGC ATGCCGCCAT GCCTGGCTAA TTTTTGGGTA TTTTTACTAG   16302
AGACAGGGTT TTGGCATGTT GTCCAGGCTG GTCTTGGACT CCTGATCTCA GATGATCCTC   16362
CTGGCTCGGG CTCCCAAAGT GCTGGGATTA CAGGCATGAA CCACCACATA TGGCCTAAAA   16422
ATTGATTCTT ATGATTAATC TCCTGTGAAC AATTTGGCTT CATTTGAAAG TTTGCCTTCA   16482
TTTGAAACCT TCATTTAAAA GCCTGAGCAA CAAAGTGAGA CCCCATCTCT ACAAAAAACT   16542
GCAAATATC CTGTGGACAC CTCCTACCTT CTGTGGAGGC TGAAGTAGGA GGATCACTTG   16602
AGCCTAGGAA TTTGAGCCTG CAGTGAGCTA TGATCCCACC TCTACTTCC AGCCTGCATG   16662
ACAGTAGACC CTGACACACA CACACAAAAA AAAACCTTCA TAAAAAATTA TTAGTTGACT   16722
TTTCTTAGGT GACTTTCCGT TTAAGCAATA AATTAAAAGA TAAAATCTCT AATTTTAGAA   16782
AATTTATTTT TAGTTACATA TTGAAATTTT TAAACCCTAG GTTTAAGTTT TATGTCTAAA   16842
TTACCTGAGA ACACACTAAG TCTGATAAGC TTCATTTTAT GGGCCTTTTG GATGATTATA   16902
TAATATTCTG ATGAAAGCCA AGACAGACCC TTAAACCATA AAAATAGGAG TTCGAGAAAG   16962
AGGAGTAGCA AAAGTAAAAG CTAGAATGAG ATTGAATTCT GAGTCGAAAT ACAAAATTTT   17022
```

- 50 -

ACATATTCTG TTTCTCTCTT TTTCCCCCTC TTAG CT GAA GAT GAT G   GTAAAGT   17075
                                          Ala Glu Asp Asp Glu
                                          -10

AGAAATGAAT TTATTTTTCT TTGCAAACTA AGTATCTGCT TGAGACACAT CTATCTCACC   17135
ATTGTCAGCT GAGGAAAAAA AAAATGGTT CTCATGCTAC CAATCTGCCT TCAAGAAAT   17195
GTGGACTCAG TAGCACAGCT TTGGAATGAA GATGATCATA AGAGATACAA AGAAGAACCT   17255
CTAGCAAAAG ATGCTTCTCT ATGCCTTAAA AAATTCTCCA GCTCTTAGAA TCTACAAAAT   17315
AGACTTTGCC TGTTTCATTG GTCCTAAGAT TAGCATGAAG CCATGGATTC TGTTGTAGGG   17375
GGAGCGTTGC ATAGGAAAAA GGGATTGAAG CATTAGAATT GTCCAAAATC AGTAACACCT   17435
CCTCTCAGAA ATGCTTTGGG AAGAAGCCTG GAAGGTTCCG GGTTGGTGGT GGGGTGGGGC   17495
AGAAAATTCT GGAAGTAGAG GAGATAGGAA TGGGTGGGGC AAGAAGACCA CATTCAGAGG   17555
CCAAAAGCTG AAAGAAACCA TGGCATTTAT GATGAATTCA GGGTAATTCA GAATGGAAGT   17615
AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGAT AAACAGGGTG TAGAGCAAGA   17675
CGTTCTCTCA CCCCAAGATG TGAAATTTGG ACTTTATCTT GGAGATAATA GGGTTAATTA   17735
AGCACAATAT GTATTAGCTA GGGTAAAGAT TAGTTTGTTG TAACAAAGAC ATCCAAAGAT   17795
ACAGTAGCTG AATAAGATAG AGAATTTTTC TCTCAAAGAA AGTCTAAGTA GGCAGCTCAG   17855
AAGTAGTATG GCTGGAAGCA ACCTGATGAT ATTGGGACCC CCAACCTTCT TCAGTCTTGT   17915
ACCCATCATC CCCTAGTTGT TGATCTCACT CACATAGTTG AAAATCATCA TACTTCCTGG   17975
GTTCATATCC CAGTTATCGA GAAAGGGTCA AGAGAAGTCA GGCTCATTCG TTTCAAAGAC   18035
TCTAATTGGA AGTTAAACAC ATCAATCCCC CTCATATTCC ATTGACTAGA ATTTAATCAC   18095
ATGGCCACAC CAAGTGCAAG GAAATCTGGA AAATATAATC TTTATTCCAG GTAGCCATAT   18155
GACTCTTTAA AATTCAGAAA TAATATATTT TTAAAATATC ATTCTGGCTT TGGTATAAAG   18215
AATTGATGGT GTGGGGTGAG GAGGCCAAAA TTAAGGGTTG AGAGCCTATT ATTTTAGTTA   18275
TTACAAGAAA TGATGGTGTC ATGAATTAAG GTAGACATAG GGGAGTGCTG ATGAGGAGCT   18335
GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAGGAC ATGATTTAGG GTTGGATTTG   18395
GAAAGGAGAA GAAAGTAGAA AAGATGATGC CTACATTTTT CACTTAGGCA ATTTGTACCA   19455
TTCAGTGAAA TAGGGAACAC AGGAGGAAGA GCAGGTTTTG GTGTATACAA AGAGGAGGAT   18515
GGATGACGCA TTTCGTTTTG GATCTGAGAT GTCTGTGGAA CGTCCTAGTG GAGATGTCCA   18575
CAAACTCTTC TACATGTGGT TCTGAGTTCA GGACACAGAT TTGGGCTGGA GATAGAGATA   18635
TTGTAGGCTT ATACATAGAA ATGGCATTTG AATCTATAGA GATAAAAAGA CACATCAGAG   18695
GAAATGTGTA AAGTGAGAGA GGAAAAGGCA AGTACTGTGC TGGGGGGAAT ACCTACATTT   18755
AAAGGATGCA GTAGAAAGAA GCTAATAAAC AACAGAGAGC AGACTAACCA AAAGGGGAGA   18815
AGAAAAACCA AGAGAATTCC ACCGACTCCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG   18875
TGTTGTGTTG AATTTTGCAG CCTTGAGAAT CAAGGGCCAG AACACAGCTT TTAGATTTAG   18935
CAACAAGGAG TTTGGTGATC TCAGTGAAAG CAGCTTGATG GTGAAATGGA GGCAGAGGCA   18995
GATTGCAATG AGTGAAACAG TGAATGGGAA GTGAAGAAAT GATACAGATA ATTCTTGCTA   19055
AAAGCTTGGC TGTTAAAAGG AGGAGAGAAA CAAGACTAGC TGCAAAGTGA GATTGGGTTG   19115
ATGGAGCAGT TTTAAATCTC AAAATAAAGA GCTTTGTGCT TTTTTGATTA TGAAAATAAT   19175
GTGTTAATTG TAACTAATTG AGGCAATGAA AAAAGATAAT AATATGAAAG ATAAAAATAT   19235
AAAAACCACC CAGAAATAAT GATAGCTACC ATTTTGATAC AATATTTCTA CACTCCTTTC   19295
TATGTATATA TACAGACACA GAAATGTTTA TATTTTTATT AAAAGGGATT GTACTATACC   19355
TAAGCTGCTT TTTCTAGTTA GTGATATATA TGGACATCTC TCCATGGCAA CGAGTAATTG   19415
CAGTTATATT AAGTTCATGA TATTTCACAA TAAGGGCATA TCTTTGCCCT TTTTATTTAA   19475
TCAATTCTTA ATTGGTGAAT GTTTGTTTCC AGTTTGTTGT TGTTATTAAC AATGTTCCCA   19535
TAAGCATTCC TGTACACCAA TGTTCACACA TTTGTCTGAT TTTTTCTTCA GGATAAAACC   19595
CAGGAGGTAG AATTGCTGGG TTGATAGAAG AGAAAGGATG ATTGCCAAAT TAAAGCTTCA   19655
GTAGAGGGTA CATGCCGAGC ATAAATGGGA TCAGCCCTAG ATACCAGAAA TGGCACTTTC   19715
TCATTTCCCC TTGGGACAAA AGGGAGAGAG GCAATAACTG TGCTGCCAGA GTTAAATTTG   19775
TACGTGGAGT AGCAGGAAAT CATTTGCTGA AAATGAAAAC AGAGATGATG TTGTAGAGGT   19835
CCTGAAGAGA GCAAAGAAAA TTTGAAATTG CGGCTATCAG CTATGGAAGA GAGTGCTGAA   19395
CTGGAAAACA AAAGAAGTAT TGACAATTGG TATGCTTGTA ATGGCACCGA TTTGAACGCT   19955
TGTGCCATTG TTCACCAGCA GCCAAGTTTG GAGTTTTGTA GCAGAAAGAC GCAGAAAGAC   20015
AAATAAGTTA GGGATTTAAT ATCCTGGCCA AATGGTAGAC AAAATGAACT CTGAGATCCA   20075
GCTGCACAGG GAAGGAAGGG AAGACGGGAA GAGGTTAGAT AGGAAATACA AGAGTCAGGA   20135
GACTGGAAGA TGTTGTGATA TTTAAGAACA CATAGAGTTG GAGTAAAAGT GTAAGAAAAC   20195
TAGAAGGGTA AGAGACGGGT CAGAAAGTAG GCTATTTGAA GTTAACACTT CAGAGGCAGA   20255
GTAGTTCTGA ATGGTAACAA GAAATTGAGT GTGCCTTTGA GAGTAGGTTA AAAAACAATA   20315
GGCAACTTTA TTGTAGCTAC TTCTGGAACA GAAGATTGTC ATTAATAGTT TTAGAAAACT   20375
AAAATATATA GCATACTTAT TTGTCAATTA ACAAAGAAAC TATGTATTTT TAAATGAGAT   20435
TTAATGTTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA   20486
                     Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu
                     -5   1   5

TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT   20534
Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile
                10   15   20

GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT   20582
Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys

```
              25                    30                      35
AGA G    GT ATTTTTTTA ATTCGCAAAC ATAGAAATGA CTAGCTACTT CTTCCCATTC  20633
Arg Asp
     40
TGTTTTACTG CTTACATTGT TCCGTGCTAG TCCCAATCCT CAGATGAAAA GTCACAGGAG  20693
TGACAATAAT TTCACTTACA GGAAACTTTA TAAGGCATCC ACGTTTTTTA GTTGGGGTAA  20758
AAAATTGGAT ACAATAAGAC ATTGCTAGGG GTCATGCCTC TCTGAGCCTG CCTTTGAATC  20818
ACCAATCCCT TTATTGTGAT TGCATTAACT GTTTAAAACC TCTATAGTTG GATGCTTAAT  20878
CCCTGCTTGT TACAGCTGAA AATGCTGATA GTTTACCAGG TGTGGTGGCA TCTATCTGTA  20938
ATCCTAGCTA CTTGGGAGGC TCAAGCAGGA GGATTGCTTG AGGCCAGGAC TTTGAGGCTG  20998
TAGTACACTG TGATCGTACC TGTGAATAGC CACTGCACTC CAGCCTGGGT GATATACAGA  21058
CCTTGTCTCT AAAATTAAAA AAAAAAAAAA AAAAACCTT AGGAAAGGAA ATTGATCAAG  21118
TCTACTGTGC CTTCCAAAAC ATGAATTCCA AATATCAAAG TTAGGCTGAG TTGAAGCAGT  21178
GAATGTGCAT TCTTTAAAAA TACTGAATAC TTACCTTAAC ATATATTTTA AATATTTTAT  21238
TTAGCATTTA AAAGTTAAAA ACAATCTTTT AGAATTCATA TCTTTAAAAT ACTCAAAAAA  21298
GTTGCAGCGT GTGTGTTGTA ATACACATTA AACTGTGGGG TTGTTTGTTT GTTTGAGATG  21358
CAGTTTCACT CTGTCACCCA GGCTGAAGTG CAGTGCAGTG CAGTGGTGTG ATCTCGGCTC  21418
ACTACAACCT CCACCTCCCA CGTTCAAGCC ATTCTCATGC CTCAGTCTCC CGAGTAGGTG  21478
GGATTACAGG CATGCACCAC TTACACCCGG CTAATTTTTG TATTTTTAGT AGAGCTGGGG  21538
TTTCACCATG TTGGCCAGGC TGGTCTCAAA CCCCTAACCT CAAGTGATCT GCCTGCCTCA  21598
GCCTCCCAAA CAAACAAACA ACCCACAGT TTAATATGTG TTACAACACA CATGCTGCAA  21658
CTTTTATGAG TATTTTAATG ATATAGATTA TAAAAGGTTG TTTTTAACTT TTAAATGCTG  21718
GGATTACAGG CATGAGCCAC TGTGCCAGGC CTGAACTGTG TTTTTAAAAA TGTCTGACCA  21778
GCTGTACATA GTCTCCTGCA GACTGGCCAA GTCTCAAAGT GGGAACAGGT GTATTAAGGA  21838
CTATCCTTTG GTTAAATTTC CGCAAATGTT CCTGTGCAAG AATTCTTCTA ACTAGAGTTC  21898
TCATTTATTA TATTTATTTC AG  AT AAT GCA CCC CGG ACC ATA TTT ATT ATA  21949
                        Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile
                        40                     45
AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT  21997
Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile Ser
50                   55                   60                   65
GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT  22045
Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile
                        70                   75                   80
TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTTCAAT CATGTTAATA TAATCAATAT  22103
Ser Phe Lys
AATTAGAAAT ATAACATTAT TTCTAATGTT AATATAAGTA ATGTAATTAG AAAACTCAAA  22163
TATCCTCAGA CCAACCTTTT GTCTAGAACA GAAATAACAA GAAGCAGAGA ACCATTAAAG  22223
TGAATACTTA CTAAAAATTA TCAAACTCTT TACCTATTGT GATAATGATG GTTTTTCTGA  22283
GCCTGTCACA GGGGAAGAGG AGATACAACA CTTGTTTTAT GACCTGCATC TCCTGAACAA  22323
TCAGTCTTTA TACAAATAAT AATGTAGAAT ACATATGTGA GTTATACATT TAAGAATAAC  22403
ATGTGACTTT CCAGAATCAG TTCTGCTATG AAGAATGAAG CTAATTATCC TTCTATATTT  22463
CTACACCTTT GTAAATTATG ATAATATTTT AATCCCTAGT TGTTTTGTTG CTGATCCTTA  22523
GCCTAAGTCT TAGACACAAG CTTCAGCTTC CAGTTGATGT ATGTTATTTT TAATGTTAAT  22583
CTAATTGAAT AAAAGTTATG AGATCAGCTG TAAAAGTAAT GCTATAATTA TCTTCAAGCC  22643
AGGTATAAAG TATTTCTGGC CTCTACTTTT TCTCTATTAT TCTCCATTAT TATTCTCTAT  22703
TATTTTTCTC TATTTCCTC ATTATTGTTA GATAAACCAC AATTAACTAT AGTTACAGAC  22763
TGAGCCAGTA AGAGTAGCCA GGGATGCTTA CAAATTGGCA ATGCTTCAGA GGAGAATTCC  22823
ATGTCATGAA GACTCTTTTT GAGTGGAGAT TTGCCAATAA ATATCCGCTT TCATGCCCAC  22883
CCAGTCCCCA CTGAAAGACA GTTAGGATAT GACCTTAGTG AAGGTACCAA GGGGCAACTT  22943
GGTAGGGAGA AAAAAGCCAC TCTAAAATAT AATCCAAGTA AGAACAGTGC ATATGCAACA  23003
GATACAGCCC CCAGACAAAT CCCTCAGCTA TCTCCCTCCA ACCAGAGTGC CACCCCTTCA  23063
GGTGACAATT TGGAGTCCCC ATTCTAGACC TGACAGGCAG CTTAGTTATC AAAATAGCAT  23123
AAGAGGCCTG GGATGGAAGG GTAGGGTGGA AAGGGTTAAG CATGCTGTTA CTGAACAACA  23183
TAATTAGAAG GGAAGGAGAT GGCCAAGCTC AAGCTATGTG GGATAGAGGA AAACTCAGCT  23243
GCAGAGGCAG ATTCAGAAAC TGGGATAAGT CCGAACCTAC AGGTGGATTC TTGTTGAGGG  23303
AGACTGGTGA AAATGTTAAG AAGATGGAAA TAATGCTTGG CACTTAGTAG GAACTGGGCA  23363
AATCCATATT TGGGGGAGCC TGAAGTTTGA TCAATTTTGA TGGCCCTTTT AAATAAAAAG  23423
AATGTGGCTG GGCGTGGTGG CTCACACCTG TAATCCCAGC ACTTTGGGAG GCCGAGGCGG  23483
GCGGATCACC TGAAGTCAGG AGTTCAAGAC CAGCCTGACC AACATGGAGA AACCCCATCT  23543
CTACTAAAAA TACAAAATTA GCTGGGCGTG GTGGCATATG CCTGTAATCC CAGCTACTCG  23603
GGAGGCTGAG GCAGGAGAAT CTTTTGAACC CGGGAGGCAG AGGTTGCGAT GAGCCTAGAT  23663
CGTGCCATTG CACTCCAGCC TGGGCAACAA GAGCAAAACT CGGTCTCAAA AAAAAAAAAA  23723
AAAAAGTGAA ATTAACCAAA GGCATTAGCT TAATAATTTA ATACTGTTTT TAAGTAGGGC  23783
GGGGGGTGGC TGGAAGAGAT CTGTGTAAAT GAGGAATCT GACATTTAAG CTTCATCAGC  23843
ATCATAGCAA ATCTGCTTCT GGAAGGAACT CAATAAATAT TAGTTGGAGG GGGGAGAGA  23903
GTGAGGGGTG GACTAGGACC AGTTTTAGCC CTTGTCTTTA ATCCCTTTTC CTGCCACTAA  23963
```

- 52 -

```
TAAGGATCTT AGCAGTGGTT ATAAAAGTGG CCTAGGTTCT AGATAATAAG ATACAACAGG  24023
CCAGGCACAG TGGCTCATGC CTATAATCCC AGCACTTTGG GAGGCCAAGG CGAGTGTCTC  24083
ACTTGAGATC AGGAGTTCAA GACCAGCCTG GCCAGCATGG CGATACTCTG TCTCTACTAA  24143
AAAAAATACA AAAATTACAT AGGCATGGTG GCATGCACCT GTAATCCCAG CTACTCGTGA  24203
GCCTGAGGCA GAAGAATCGC TTGAAACCAG GAGGCGTAGG CTGCAGTGAG CTGAGATCGC  24263
ACCACTGCAC TCCAGCCTGG GCGACAGAAT GAGACTTTGT CTCAAAAAAA GAAAAAGATA  24323
CAACAGGCTA CCCTTATGTG CTCACCTTTC ACTGTTGATT ACTAGCTATA AAGTCCTATA  24383
AAGTTCTTTG GTCAAGAACC TTGACAACAC TAAGAGGGAT TTGCTTTGAG AGGTTACTGT  24443
CAGAGTCTGT TTCATATATA TACATATACA TGTATATATG TATCTATATC CAGGCTTGGC  24503
CAGGGTTCCC TCAGACTTTC CAGTGCACTT GGGAGATGTT AGGTCAATAT CAACTTTCCC  24563
TGGATTCAGA TTCAACCCCT TCTGATGTAA AAAAAAAAAA AAAAAGAAA GAAATCCCTT  24623
TCCCCTTGGA GCACTCAAGT TTCACCAGGT GGGGCTTTCC AAGTTGGGGG TTCTCCAAGG  24683
TCATTGGGAT TGCTTTCACA TCCATTTGCT ATGTACCTTC CCTATGATGG CTGGGAGTGG  24743
TCAACATCAA AACTAGGAAA GCTACTGCCC AAGGATGTCC TTACCTCTAT TCTGAAATGT  24803
GCAATAAGTG TGATTAAAGA GATTGCCTGT TCTACCTATC CACACTCTCG CTTTCAACTG  24863
TAACTTTCTT TTTTTCTTTT TTTCTTTTTT TCTTTTTTTT TGAAACGGAG TCTCGCTCTG  24923
TCGCCCAGGC TAGAGTGCAG TGGCACGATC TCAGCTCACT GCAAGCTCTG CCTCCCGGGT  24983
TCACGCCATT CTCCTGCCTC ACCCTCCCAA GCAGCTGGGA CTACAGGCGC CTGCCACCAT  25043
GCCCAGCTAA TTTTTTGTAT TTTTAGTAGA GACGGGGTTT CACCGTGTTA GCCAGGATGG  25103
TCTCGATCTC CTGAACTTGT GATCCGCCCG CCTCAGCCTC CCAAAGTGCT GGGATTACAG  25163
GCGTGAGCCA TCGCACCCGG CTCAACTGTA ACTTTCTATA CTGGTTCATC TTCCCCTGTA  25223
ATGTTACTAG AGCTTTTGAA GTTTTGGCTA TGGATTATTT CTCATTTATA CATTAGATTT  25283
CAGATTAGTT CCAAATTGAT GCCCACAGCT TAGGGTCTCT TCCTAAATTG TATATTGTAG  25343
ACAGCTGCAG AAGTGGGTGC CAATAGGGGA ACTAGTTTAT ACTTTCATCA ACTTAGGACC  25403
CACACTTGTT GATAAAGAAC AAAGGTCAAG AGTTATGACT ACTGATTCCA CAACTGATTG  25463
AGAAGTTGGA GATAACCCCG TGACCTCTGC CATCCAGAGT CTTTCAGGCA TCTTTGAAGG  25523
ATGAAGAAAT GCTATTTTAA TTTTGGAGGT TTCTCTATCA GTGCTTAGGA TCATGGGAAT  25583
CTGTGCTGCC ATGAGGCCAA AATTAACATC AAAACATCCA CTGGTTCCAG GATTAACATG  25643
GAAGAACCTT AGGTGGTGCC CACATGTTCT GATCCATCCT GCAAAATAGA CATGCTGCAC  25703
TAACAGGAAA AGTGCAGGCA GCACTACCAG TTGGATAACC TGCAAGATTA TAGTTTCAAG  25763
TAATCTAACC ATTTCTCACA AGGCCCTATT CTGTGACTGA AACATACAAG AATCTGCATT  25823
TGGCCTTCTA AGGCAGGGCC CAGCCAAGGA GACCATATTC AGGACAGAAA TTCAAGACTA  25883
CTATGGAACT GGAGTGCTTG GCAGGGAAGA CAGAGTCAAG GACTGCCAAC TGAGCCAATA  25943
CAGCAGGCTT ACACAGGAAC CCAGGGCCTA GCCCTACAAC AATTATTGGG TCTATTCACT  26003
GTAAGTTTTA ATTTCAGGCT CCACTGAAAG AGTAAGCTAA GATTCCTGGC ACTTTCTGTC  26063
TCTCTCACAG TTGGCTCAGA AATGAGAACT GGTCAGGCCA GGCATGGTGG CTTACACCTG  26123
GAATCCCAGC ACTTTGGGAG GCCGAAGTGG GAGGGTCACT TGAGGCCAGG AGTTCAGGAC  26183
CAGCTTAGGC AACAAAGTGA GATACCCCCT GATCCCTTCT CTACAAAAAT AAATTTTAAA  26243
AATTAGCCAA ATGTGGTGGT GTATACTTAC AGTCCCAGCT ATTCAGGAGG CTGAGGCAGG  26303
GGGATTGCTT GAGCCCAGGA ATTCAAGGCT GCAGTGAGCT ATGATTTCAC CACTGCACTT  26363
CTGGCTGGGC AACAGAGCGA GACCCTGTCT CAAAGCAAAA AGAAAAAGAA ACTAGAACTA  26423
GCCTAAGTTT GTGGGAGGAG GTCATCATCG TCTTTAGCCG TGAATGGTTA TTATAGAGGA  26483
CAGAAATTGA CATTAGCCCA AAAAGCTTGT GGTCTTTGCT GGAACTCTAC TTAATCTTGA  26543
GCAAATGTGG ACACCACTCA ATGGGAGAGG AGAGAAGTAA GCTGTTTGAT GTATAGGGGA  26603
AAACTAGAGG CCTGGAACTG AATATGCATC CCATGACAGG GAGAATAGGA GATTCGGAGT  26663
TAAGAAGGAG AGGAGGTCAG TACTGCTGTT CAGAGATTTT TTTTATGTAA CTCTTGAGAA  26723
GCAAAACTAC TTTTGTTCTG TTTGGTAATA TACTTCAAAA CAAACTTCAT ATATTCAAAT  26783
TGTTCATGTC CTGAAATAAT TAGGTAATGT TTTTTTCTCT ATAG GAA ATG AAT CCT  26839
                                                  Glu Met Asn Pro
                                                       85

CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA    26887
Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg
     90            95                100

AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC    26935
Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr
105          110             115                120

GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC    26983
Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu
             125             130                135

ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT    27031
Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr
             140             145                150

GTT CAA AAC GAA GAC T AGCTATTAAA ATTTCATGCC GGGCGCAGTG GCTCACGCCT   27087
Val Gln Asn Glu Asp
     155

GTAATCCCAG CCCTTTGGGA GGCTGAGGCG GGCAGATCAC CAGAGGTCAG GTGTTCAAGA   27147
CCAGCCTGAC CAACATGGTG AAACCTCATC TCTACTAAAA ATACAAAAAA TTAGCTGAGT   27207
```

- 53 -

```
GTAGTGACCC ATGCCCTCAA TCCCAGCTAC TCAAGAGGCT GAGGCAGGAG AATCACTTGC 27267
ACTCCGGAGG TGGAGGTTGT GGTGAGCCGA GATTGCACCA TTGCGCTCTA GCCTGGGCAA 27327
CAACAGCAAA ACTCCATCTC AAAAAATAAA ATAAATAAAT AAACAAATAA AAAATTCATA 27387
ATGTGAACTG TCTGAATTTT TATGTTTAGA AAGATTATGA GATTATTAGT CTATAATTGT 27447
AATGGTGAAA TAAAATAAAT ACCAGTCTTG AAAAACATCA TTAAGAAATG AATGAACTTT 27507
CACAAAAGCA AACAAACAGA CTTTCCCTTA TTTAAGTGAA TAAAATAAAA TAAAATAAAA 27567
TAATGTTTAA AAAATTCATA GTTTGAAAAC ATTCTACATT GTTAATTGGC ATATTAATTA 27627
TACTTAATAT AATTATTTTT AAATCTTTTG GGTTATTAGT CCTAATGACA AAAGATATTG 27687
ATATTTGAAC TTTCTAATTT TTAAGAATAT CGTTAAACCA TCAATATTTT TATAAGGAGG 27747
CCACTTCACT TGACAAATTT CTGAATTTCC TCCAAAGTCA GTATATTTTT AAAATTCAGT 27807
TTGATCCTGA ATCCAGCAAT ATATAAAAGG GATTATATAC TCTGGCCAAC TGACATTCAT 27867
CCTAGGAATG CAAAGATGGT TTAATATCCT AAAATCAATT AACATAACAT ACTATATTAA 27927
TAAAGTATCA AAACAGTATT CTCATCTTTT TTTCTTTTTT CACAATTCCT TGGTTACACT 27987
ATCATCTCAA TAGATGCAGA AAAAGCATTT GACAAAATCC AATTCATAAT AAAAATTCTC 28047
AAACTTGAAA GAGAACATCA TAAAGGCATC TATGAAAAAC CTACAGCTAA TATCATACTT 28107
AACGATGAAA AACTGAATTA TTTTACCCTA AGATCAAGAA TAATGCAAGC ATGTCAGCTC 28167
TTGCAACTTC TATTCAACAT TGTACTGGAG GTTCTAGCCA GAGCAACCAT ACAATAAATA 28227
AAAATAAAAG GCATCCAGAT TAGAAAGGAA GTCTTTATTT GCAGACAACA TGGTTCTTTA 28287
TGCAGAAAAC CGTCAGGAAT ACACACACAT GTTAGAACTA ATAAGTTCAG CAAGGTTGCA 28347
GGTTGCAATA TCAATATGCA AAAATACATT GAAGGCTGGG CTCAGTGGAG ATGGCATGTA 28407
CCTTTCGTCC CAGCTACTTG GGAGGCTGAG GTAGGAGGAT CACTTGAGGT GAGGAGTTTG 28467
AGGCTATAGT GCAATGTGAT CTTGCCTGTG AATAGCCACT GCACTCGAGC CTAGGCAACA 28527
AAGTGAGACC CCGTCTCCAA AAAAAAAAT GGTATATTGG TATTTCTGTA TATGAACAAT 28587
GAATGATCTG AAAACAAGAA AATTCCATTC ACGATGGTAT TAAAAAAATA AAATACAAAT 28647
AAATTTAGCA AAATAATTAT AAAACTTGTA CATCGAAAAT TTCAAAGCAC TCTGAGGGAA 28707
ATTAAAGATG ATCTAAATAA TTGGAGAGAC ACTCTATGAT CACTGATTGG AAAATTCATT 28767
CAATATTGTT AAGATAACAA TTGTCCCCAA ATTGATGCAT GCATTCAATT TAGTCTTCAT 28827
CAAAATTCCA GCAGGGTTTT TGCAGAAATT GACAAGCTGT ACCCAAAATG TATATGGAAA 28887
TGAAAAGACC CAGAAGAGCA AATAATTTTT TAAAAACAAA GTTGGAAAAC TTTTACTTCC 28947
TAATTTTAAA ACTTACTATA AACCTAAAGT TATCAAGACC ATTTAGT         28994
```

(15) INFORMATION FOR SEQ ID NO: 15:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 10 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: peptide

      (v) FRAGMENT TYPE: N-terminal fragment

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO:16:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 27 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
CCATCCTAAT ACGACTCACT ATAGGGC                               27
```

(2) INFORMATION FOR SEQ ID NO:17:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTCCTCTTCC CGAAGCTGTG TAGACTGC   28

(2) INFORMATION FOR SEQ ID NO:18:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 18 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTATAGGGCA CGCGTGGT   18

(2) INFORMATION FOR SEQ ID NO:19:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 28 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCCTCTTCC CGAAGCTGTG TAGACTGC   28

(2) INFORMATION FOR SEQ ID NO:20:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 30 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAGTTTTC ACCTTCCAAC TGTAGAGTCC   30

(2) INFORMATION FOR SEQ ID NO:21:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 30 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGGATCAAGT CGTGATCAGA AGCAGCACAC   30

(2) INFORMATION FOR SEQ ID NO:22:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCTGGCTGCC AACTCTGGCT GCTAAAGCGG   30

(2) INFORMATION FOR SEQ ID NO:23:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 33 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTATTGTCAA TAAATTTCAT TGCCACAAAG TTG   33

(2) INFORMATION FOR SEQ ID NO:24:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 33 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAGATGGCTG CTGAACCAGT AGAAGACAAT TGC   33

(2) INFORMATION FOR SEQ ID NO:25:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 27 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCCTTGGTCA ATGAAGAGAA CTTGGTC   27

(2) INFORMATION FOR SEQ ID NO:26:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC   33

(2) INFORMATION FOR SEQ ID NO:27:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 32 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG   32

(2) INFORMATION FOR SEQ ID NO:28:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 31 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCAGCCTAG AGGTATGGCT GTAACTATCT C   31

(2) INFORMATION FOR SEQ ID NO:29:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 33 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATGAAAT TTTAATAGCT AGTCTTCGTT TTG   33

(2) INFORMATION FOR SEQ ID NO:30:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 30 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGACATCAT ATTCTTTCAG AGAAGTGTCC   30

(2) INFORMATION FOR SEQ ID NO:31:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 31 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAATTTGAA TCTTCATCAT ACGAAGGATA C   31

(2) INFORMATION FOR SEQ ID NO:32:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 31 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCCGAAGCTT AAGATGGCTG CTGAACCAGT A   31

(2) INFORMATION FOR SEQ ID NO:33:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 32 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG   32

(2) INFORMATION FOR SEQ ID NO:34:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 37 base pairs
             (B) TYPE: nucleic acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear

      (ii) MOLECULE TYPE: cDNA

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ATGTAGCGGC CGCGGCATGA AATTTTAATA GCTAGTC   37

(2) INFORMATION FOR SEQ ID NO:35:

       (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC   33